

CC protein. The method comprises providing a sample of a body fluid and/or
 CC tissue of a patient and determining a level of the aberrant protein
 CC and/or its detectable part or the ratio between the two in the body fluid
 CC and/or tissue. The method enables diagnosing the disease before the
 CC patient exhibits clinical symptoms. The method is useful for diagnosing
 CC age-related diseases, preferably Alzheimer's disease, Down's syndrome or
 CC other age related diseases e.g. cancer; neurodegenerative diseases, such
 CC as frontal lobe dementia, progressive supranuclear palsy; and other
 CC diseases with abundant tau-positive filamentous lesions, Parkinson's
 CC disease, amyotrophic lateral sclerosis, Huntington's disease,
 CC spinocerebellar ataxia-3, multiple sclerosis; other inclusion body
 CC diseases associated with ubiquitin, diabetes mellitus type II and other
 CC degenerative diseases such as cardiovascular diseases and rheumatoid
 CC arthritis. A number of other diseases which can be diagnosed are given in
 CC the specification. The present sequence is AMY 6 peptide, which was
 CC coupled to thyroglobulin by glutaraldehyde, and used to immunise
 CC rabbits for antibody production in the present invention. The AMY 6
 CC peptide is a C-terminal fragment of the full-length beta-amyloid
 CC precursor protein (APP+1; mature APP and its signal peptide) protein.
 XX
 SQ Sequence 19 AA;

Query Match 83.3%; Score 82.5; DB 22; Length 19;
 Best Local Similarity 94.7%; Pred. No. 1.6e-07;
 Matches 18; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 NVPCHERMGRITSSKELA 19
 DB 2 NVP-HERMGRITSSKELA 19
 ||||| ||||| ||||| ||||| |||||

RESULT 8
 AAB99215
 ID AAB99215 standard; peptide; 14 AA.

XX
 AC AAB99215;

XX
 DT 06-SEP-2001 (first entry)

XX
 DE AMY5 peptide.

XX AMY5; disease detection; age-related disease; Alzheimer's disease;
 KW Down's syndrome; cancer; neurodegenerative disease; Parkinson's disease;
 KW amyotrophic lateral sclerosis; Huntington's disease;
 KW spinocerebellar ataxia-3; multiple sclerosis; diabetes mellitus type II;
 KW degenerative disease; cardiovascular disease; rheumatoid arthritis;
 KW beta-amyloid precursor protein; APP.

XX Unidentified.

XX WO200140804-A2.

XX
 PD 07-JUN-2001.

XX
 PF 04-DEC-2000; 2000WO-NL00887.

XX
 PR 03-DEC-1999; 99EP-0204140.

XX (NEW-) KONINK NEDERLANDSE AKAD VAN WETENSCHAPPE.

XX
 PI Hol EM, van Leeuwen FW;

XX
 DR WPI; 2001-397965/42.

XX
 PT Detecting a disease or pathological condition, associated with
 PT secretion of aberrant protein e.g. age-related diseases, by determining
 PT level of aberrant protein and/or its detectable part in body fluid
 PT and/or tissue

XX
 PS Disclosure; Page 24; 36pp; English.

XX
 CC The present invention relates to a method for detecting a disease or
 CC pathological condition associated with molecular misreading of coding

CC sequences in the genome and/or associated with the disease, and/or associated
 CC protein. The method comprises providing a sample of a body fluid and/or
 CC tissue of a patient and determining a level of the aberrant protein
 CC and/or its detectable part or the ratio between the two in the body fluid
 CC and/or tissue. The method enables diagnosing the disease before the
 CC patient exhibits clinical symptoms. The method is useful for diagnosing
 CC age-related diseases, preferably Alzheimer's disease, Down's syndrome or
 CC other age related diseases e.g. cancer; neurodegenerative diseases, such
 CC as frontal lobe dementia, progressive supranuclear palsy; and other
 CC diseases with abundant tau-positive filamentous lesions, Parkinson's
 CC disease, amyotrophic lateral sclerosis, Huntington's disease,
 CC spinocerebellar ataxia-3, multiple sclerosis; other inclusion body
 CC diseases associated with ubiquitin, diabetes mellitus type II and other
 CC degenerative diseases such as cardiovascular diseases and rheumatoid
 CC arthritis. A number of other diseases which can be diagnosed are given in
 CC the specification. The present sequence is AMY 5 peptide, which was
 CC coupled to thyroglobulin by glutaraldehyde, and used to immunise
 CC rabbits for antibody production in the present invention. The AMY 5
 CC peptide is a C-terminal fragment of the full-length beta-amyloid
 CC precursor protein (APP+1; mature APP and its signal peptide) protein.
 XX
 SQ Sequence 14 AA;

Query Match 63.6%; Score 63; DB 22; Length 14;
 Best Local Similarity 100.0%; Pred. No. 0.00033;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 RMGRGRITSSKELA 19
 DB 2 RMGRGRITSSKELA 14
 ||||| ||||| ||||| |||||

RESULT 9
 AAY21522
 ID AAY21522 standard; Protein; 28 AA.

XX
 AC AAY21522;

XX
 DT 22-JUL-1999 (first entry)

XX
 DE Human beta-APP exon 10 protein fragment.

XX Human; beta-amyloid precursor protein; beta-APP; diagnosis; cancer;
 KW frameshift mutation; age-related disease; neurodegenerative disorder;
 KW Alzheimer's disease; Down's syndrome; myotonic dystrophy; neuronal;
 KW Huntington's disease; multiple sclerosis; alcoholic liver disease;
 KW diabetes mellitus type II; microtubule associated protein; Tau; Big Tau;
 KW ubiquitin B; apolipoprotein E; MAP2; neurofilament-L; neurofilament-M;
 KW neurofilament-F; presenilin 1; presenilin 2; cellular tumour antigen;
 KW glial fibrillary acidic protein; GFAP; p53; semaphorin III; HUPF-1;
 KW bcl-2; B-cell leukemia/lymphoma 2 proto-oncogene; HMGP-C; NSP-A;
 KW high mobility group protein-C; neuroendocrine specific protein A.

XX
 OS Synthetic.

XX
 OS Homo sapiens.

XX
 PN WO9845322-A2.

XX
 PD 15-OCT-1998.

XX
 PF 02-APR-1998; 98WO-1B00705.

XX
 PR 10-APR-1997; 97US-0043163.

XX
 PA (UYUT-) RIJKSUNIV UTRECHT.

XX
 PA (ROYA-) ROYAL NETHERLANDS ACAD ARTS & SCI.

XX
 PA (UYRO-) UNIV ROTTERDAM ERASMUS.

XX
 PI Burbach JPH, Grosveld FG, Van Leeuwen FW;

XX
 DR WPI; 1998-609901/51.

XX
 DR N-PSDB; AAX75772.

Query Match 52.5%; Score 52; DB 21; Length 17;
 Best Local Similarity 100.0%; Pred. No. 0.036;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NVPGRMG 9
 | | | | | | | | |
 DB 9 NVPGRMG 17

RESULT 12
 AAJ46738
 ID AAU46738 standard; Protein; 57 AA.
 XX AC AAU46738;
 XX DT 27-FEB-2002 (first entry)
 XX DE Propionibacterium acnes immunogenic protein #7634.
 XX KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.
 XX OS Propionibacterium acnes.
 XX PN WO200181581-A2.
 XX PD 01-NOV-2001.
 XX PF 20-APR-2001; 2001WO-US12865.
 XX PR 21-APR-2000; 2000US-199047P.
 PR 02-JUN-2000; 2000US-208841P.
 PR 07-JUL-2000; 2000US-216747P.
 XX PA (CORI-) CORIXA CORP.
 XX PI Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
 PI L'Maisonneuve J, Zhang Y, Jen S, Carter D;
 XX WPI: 2001-616774/71.
 DR N-PSDB; AAS59535.
 XX PT Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris -
 XX Example 1; SEQ ID No 7933; 1069pp; English.
 XX CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA).
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 57 AA;

Query Match 51.5%; Score 51; DB 22; Length 57;
 Best Local Similarity 81.8%; Pred. No. 0.2;
 Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 VPGHERMGR 12
 | | | | | | | | | | | |
 DB 47 VPGHRETGR 57

RESULT 13
 AAU66292
 ID AAU66292 standard; Protein; 57 AA.
 XX AC AAU66292;
 XX DT 27-FEB-2002 (first entry)
 XX DE Propionibacterium acnes immunogenic protein #27188.
 XX KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.
 XX OS Propionibacterium acnes.
 XX PN WO200181581-A2.
 XX PD 01-NOV-2001.
 XX PF 20-APR-2001; 2001WO-US12865.
 XX PR 21-APR-2000; 2000US-199047P.
 PR 02-JUN-2000; 2000US-208841P.
 PR 07-JUL-2000; 2000US-216747P.
 XX PA (CORI-) CORIXA CORP.
 XX PI Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
 PI L'Maisonneuve J, Zhang Y, Jen S, Carter D;
 XX WPI: 2001-616774/71.
 DR N-PSDB; AAS59726.
 XX PT Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris -
 XX Example 1; SEQ ID No 27487; 1069pp; English.
 XX CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA).
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 57 AA;

```

Query Match      51.5%; Score 51; DB 22; Length 57;
Best Local Similarity 81.8%; Pred. No. 0.2;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 VPGHERMGRGR 12
Db 47 VPGHRTGGR 57

RESULT 14
ABB59130
ID ABB59130 standard; Protein: 491 AA.
XX
AC ABB59130;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 4182.
XX
KW Drosophila; development; biology; cell signalling; insecticide;
XX
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
DR N-PSDB; ABL03233.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 4182; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 491 AA:

Query Match      51.0%; Score 50.5; DB 22; Length 491;
Best Local Similarity 48.0%; Pred. No. 2.6;
Matches 12; Conservative 3; Mismatches 3; Indels 7; Gaps 1;

Qy 2 VPGHERMGRGRSS-----KELA 19
Db 171 LPSHERLGLGRSLQFTIELTKELA 195

RESULT 15
ABB53619
ID ABB53619 standard; Protein: 72 AA.

```

```

XX ABB53619;
AC
DT 16-MAY-2002 (first entry)
XX
DE Lactococcus lactis protein ydBC.
XX
KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
XX
OS Lactococcus lactis IL1403.
XX
PN FR2807446-A1.
XX
PD 12-OCT-2001.
XX
PF 11-APR-2000; 2000FR-0004630.
XX
PR 11-APR-2000; 2000FR-0004630.
XX
PA (INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX
PI Bolotin A, Sorokine A, Renault P, Ehrlich SD;
XX
DR WPI; 2002-043418/06.
XX
PT New nucleotide sequence useful in the identification of Lactococcus
PT lactis and related species -
XX
PS Claim 6; SEQ ID NO 321; 2504pp; French.
XX
CC The present invention is related to a Lactococcus lactis nucleotide
CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The
CC nucleic acid sequence is useful in the detection and/or amplification of
CC nucleic acid sequence, particularly to identify Lactococcus lactis or
CC related species. The proteins of the invention are useful for the
CC biosynthesis or biodegradation of a composition of interest. The
CC invention helps research in lactic bacteria, particularly useful in the
CC production of yogurt and cheese.
CC Note: The sequence data for this patent is based on equivalent patent
CC WO200177334 (published 18-OCT-2001) which is available in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 72 AA:

Query Match      50.5%; Score 50; DB 23; Length 72;
Best Local Similarity 60.0%; Pred. No. 0.39;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 3 PGHERMGRGRTSSKE 17
Db 46 PDHEKMGKGITLSEF 60

Search completed: October 1, 2003, 09:38:25
Job time : 84 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2003, 09:36:59 : Search time 17 Seconds
(without alignments)
47.289 Million cell updates/sec

Title: US-09-674-913a-1

Perfect score: 99

Sequence: 1 NVPGHERMGRGRTSSKELA 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA: *
1: /cgn2.6/ptodata/1/1aa/5A-COMB.pep: *
2: /cgn2.6/ptodata/1/1aa/5B-COMB.pep: *
3: /cgn2.6/ptodata/1/1aa/6A-COMB.pep: *
4: /cgn2.6/ptodata/1/1aa/6B-COMB.pep: *
5: /cgn2.6/ptodata/1/1aa/PctUS-COMB.pep: *
6: /cgn2.6/ptodata/1/1aa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	99	100.0	25	2	US-08-726-306A-61
2	99	100.0	25	2	US-08-726-306A-87
3	47	47.5	10	2	US-08-726-306A-1
4	47	47.5	87	4	US-09-107-532A-5378
5	46	46.5	26	1	US-07-942-245-321
6	46	46.5	139	4	US-09-252-991A-23482
7	46	46.5	553	4	US-09-252-991A-17429
8	45.5	46.0	537	4	US-09-134-001C-4091
9	45	45.5	496	4	US-09-252-991A-22592
10	45	45.5	582	4	US-09-252-991A-25313
11	45	45.5	940	4	US-09-252-991A-24639
12	44	44.4	266	4	US-09-252-991A-32512
13	44	44.4	341	4	US-09-252-991A-27955
14	44	44.4	417	4	US-09-252-991A-16986
15	44	44.4	530	4	US-09-252-991A-30947
16	44	44.4	656	4	US-09-252-991A-22270
17	43	43.4	119	2	US-08-581-528A-7
18	43	43.4	119	5	PCT-US94-07799-7
19	43	43.4	161	2	US-08-581-528A-6
20	43	43.4	161	5	PCT-US94-07799-6
21	43	43.4	192	4	US-09-252-991A-8164
22	43	43.4	223	4	US-09-252-991A-17072
23	43	43.4	460	4	US-09-252-991A-26877
24	42	42.4	129	1	US-08-360-914B-15
25	42	42.4	129	1	US-08-741-589A-13
26	42	42.4	293	4	US-09-252-991A-29137
27	42	42.4	294	1	US-08-362-670B-2

28 42 42.4 294 3 US-08-333-576C-2 Sequence 2, Appli
29 42 42.4 294 3 US-08-808-324-2 Sequence 2, Appli
30 42 42.4 294 5 PCT-US94-14030A-2 Sequence 2, Appli
31 42 42.4 388 1 US-08-362-670B-34 Sequence 34, Appli
32 42 42.4 388 3 US-08-333-576C-34 Sequence 34, Appli
33 42 42.4 388 3 US-08-808-324-34 Sequence 34, Appli
34 42 42.4 388 5 PCT-US94-14030A-34 Sequence 34, Appli
35 42 42.4 411 1 US-08-362-670B-28 Sequence 28, Appli
36 42 42.4 411 3 US-08-333-576C-28 Sequence 28, Appli
37 42 42.4 411 3 US-08-808-324-28 Sequence 28, Appli
38 42 42.4 411 5 PCT-US94-14030A-28 Sequence 28, Appli
39 42 42.4 418 4 US-09-648-281-12 Sequence 12, Appli
40 42 42.4 591 4 US-09-252-991A-31286 Sequence 31286, A
41 42 42.4 592 4 US-09-252-991A-18124 Sequence 18124, A
42 42 42.4 735 4 US-09-252-991A-17053 Sequence 17053, A
43 41.5 41.9 684 4 US-09-252-991A-30736 Sequence 30736, A
44 41 41.4 123 4 US-09-732-210-1370 Sequence 1370, Ap
45 41 41.4 415 4 US-09-252-991A-27669 Sequence 27669, A

ALIGNMENTS

RESULT 1
US-08-726-306A-61
* Sequence 61, Application US/08726305A
* Patent No. 5958884
* GENERAL INFORMATION:
* APPLICANT: van Leeuwen, Frederik Willem
* APPLICANT: Burbach, Johannes Peter Henri
* APPLICANT: Grosveld, Franklin G.
* TITLE OF INVENTION: DIAGNOSIS METHOD AND REAGENTS
* NUMBER OF SEQUENCES: 189
* CORRESPONDENCE ADDRESS:
* ADDRESSEE: Banner & Witcoff, Ltd.
* STREET: 1 Financial Center
* CITY: Boston
* STATE: MA
* COUNTRY: US
* ZIP: 02111
* COMPUTER READABLE FORM:
* MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
* COMPUTER: IBM PC Compatible
* OPERATING SYSTEM: PC-DOS/MS-DOS
* SOFTWARE: Wordperfect 6.1
* CURRENT APPLICATION DATA:
* APPLICATION NUMBER: US/08/726,306A
* FILING DATE: 02-Oct-1996
* PRIOR APPLICATION DATA:
* APPLICATION NUMBER: GB 95/20080.4
* FILING DATE: 02-Oct-1995
* PRIOR APPLICATION DATA:
* APPLICATION NUMBER: US 60/009,832
* FILING DATE: 01-Jan-1996
* ATTORNEY/AGENT INFORMATION:
* NAME: Williams, Ph.D., Kathleen M.
* REGISTRATION NUMBER: 34,380
* REFERENCE/DOCKET NUMBER: 96,048-A (3255/00784)
* TELECOMMUNICATION INFORMATION:
* TELEPHONE: (617) 345-9100
* TELEFAX: (617) 345-9111
* INFORMATION FOR SEQ ID NO: 61:
* SEQUENCE CHARACTERISTICS:
* LENGTH: 25 amino acids
* TYPE: amino acid
* STRANDEDNESS: single
* TOPOLOGY: unknown
* MOLECULE TYPE: protein
US-08-726-306A-61

Query Match 100.0%; Score 99; DB 2; Length 25;
Best Local Similarity 100.0%; Pred. No. 5.le-11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 NVPGHERMGRGTSSKELA 19
DB 7 NVPGHERMGRGTSSKELA 25

RESULT 2
US-08-726-306A-87
; Sequence 87, Application US/08726306A
; Patent No. 5958684
; GENERAL INFORMATION:
; APPLICANT: van Leeuwen, Frederik Willem
; APPLICANT: Burbach, Johannes Peter Henri
; APPLICANT: Grosveld, Franklin G.
; TITLE OF INVENTION: DIAGNOSIS METHOD AND REAGENTS
; NUMBER OF SEQUENCES: 189
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1 Financial Center
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,306A
; FILING DATE: 02-Oct-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 95/20080.4
; FILING DATE: 02-Oct-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/009,832
; FILING DATE: 01-Jan-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams, Ph.D., Kathleen M.
; REGISTRATION NUMBER: 34,380
; REFERENCE/DOCKET NUMBER: 96,048-A (3255/00784)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 345-9100
; TELEFAX: (617) 345-9111
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; CS-08-726-306A-1

Query Match 47.5%; Score 47; DB 2; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.032;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 RGRITSSKELA 19
DB 1 RGRITSSKELA 10

RESULT 4
US-09-107-532A-5378
; Sequence 5378, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A. Doucette-Stamm and David Bush
; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571

QY 1 NVPGHERMGRGTSSKELA 19
DB 7 NVPGHERMGRGTSSKELA 25

RESULT 3
US-08-726-306A-1
; Sequence 1, Application US/08726306A
; Patent No. 5958684
; GENERAL INFORMATION:
; APPLICANT: van Leeuwen, Frederik Willem
; APPLICANT: Burbach, Johannes Peter Henri
; APPLICANT: Grosveld, Franklin G.
; TITLE OF INVENTION: DIAGNOSIS METHOD AND REAGENTS
; NUMBER OF SEQUENCES: 189
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1 Financial Center
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,306A
; FILING DATE: 02-Oct-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 95/20080.4
; FILING DATE: 02-Oct-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/009,832
; FILING DATE: 01-Jan-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams, Ph.D., Kathleen M.
; REGISTRATION NUMBER: 34,380
; REFERENCE/DOCKET NUMBER: 96,048-A (3255/00784)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 345-9100
; TELEFAX: (617) 345-9111
; INFORMATION FOR SEQ ID NO: 87:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-726-306A-87

Query Match 100.0%; Score 99; DB 2; Length 25;
Best Local Similarity 100.0%; Pred. No. 5.1e-11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NVPGHERMGRGTSSKELA 19
DB 7 NVPGHERMGRGTSSKELA 25

RESULT 3
US-08-726-306A-1
; Sequence 1, Application US/08726306A
; Patent No. 5958684
; GENERAL INFORMATION:
; APPLICANT: van Leeuwen, Frederik Willem
; APPLICANT: Burbach, Johannes Peter Henri
; APPLICANT: Grosveld, Franklin G.
; TITLE OF INVENTION: DIAGNOSIS METHOD AND REAGENTS
; NUMBER OF SEQUENCES: 189
; CORRESPONDENCE ADDRESS:
```

```

; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Arinello, Pamela Deneko
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5067
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5378:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 87 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (8) LOCATION 1...87
; SEQUENCE DESCRIPTION: SEQ ID NO: 5378:
US-09-107-532A-5378
Query Match 47.5%; Score 47; DB 4; Length 87;
Best Local Similarity 53.3%; Pred. No. 0.41;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 PGHERMGRGRTSSKE 17
DB 61 PDHERMKGRLTLTNE 75

RESULT 5
US-07-942-245-321
; Sequence 321, Application US/07942245
; Patent No. 5639641
; GENERAL INFORMATION:
; APPLICANT: PEDERSEN, Jan T.
; APPLICANT: REARLE, Stephen M.J.
; APPLICANT: REES, Anthony R.
; APPLICANT: ROGUSKA, Michael A.
; APPLICANT: GUTILD, Braydon C.
; TITLE OF INVENTION: SURFACE RESIDUE VENEERING OF RODENT
; NUMBER OF SEQUENCES: 522
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrie, Mion., Zinn, Macpeak & Seas
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: HP 9000/700 Workstation
; OPERATING SYSTEM: UNIX
; SOFTWARE: In house
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/942,245
; FILING DATE: 09-SEP-1992
; CLASSIFICATION: 530
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 321:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 26 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-942-245-321

; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Arinello, Pamela Deneko
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5067
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5378:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 87 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (8) LOCATION 1...87
; SEQUENCE DESCRIPTION: SEQ ID NO: 5378:
US-09-107-532A-5378
Query Match 46.5%; Score 46; DB 1; Length 26;
Best Local Similarity 66.7%; Pred. No. 0.15;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 PGHERMGRGRTS 14
DB 13 PGHEKKGRSS 24

RESULT 6
US-09-252-991A-23482
; Sequence 23482, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/034,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23482
; LENGTH: 139
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23482
Query Match 46.5%; Score 46; DB 4; Length 139;
Best Local Similarity 50.0%; Pred. No. 1.1;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 VPGHERMGRGRTSSKELA 19
DB 53 VPGRAGAGRGRTARSRA 70

RESULT 7
US-09-252-991A-17429
; Sequence 17429, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/034,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17429
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17429
Query Match 46.5%; Score 46; DB 4; Length 553;
Best Local Similarity 72.7%; Pred. No. 5.3;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 PGHERMGRGRT 13
DB 1 PDQRHGRGRT 11

RESULT 8
US-09-134-001C-4091
```

```

; Sequence 4091, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134.001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4091
; LENGTH: 537
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4091

Query Match 46.0%; Score 45.5; DB 4; Length 537;
Best Local Similarity 55.0%; Pred. No. 6.3;
Matches 11; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

Qy 1 NVPGE-ERMGRGRTSSKLA 19
I:|||||I||:||:|I
Db 186 NLPGHVAMGVQSQTJLJA 205

RESULT 9
US-09-252-991A-22592
; Sequence 22592, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22592
; LENGTH: 496
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22592

Query Match 45.5%; Score 45; DB 4; Length 496;
Best Local Similarity 36.7%; Pred. No. 7.1;
Matches 11; Conservative 1; Mismatches 0; Indels 18; Gaps 1;

Qy 2 VPGHMR-----GRGRT 13
:|||||I||:|I
Db 428 LPGHMRARLSKLTIDLRVDYLRPGRGRT 457

RESULT 10
US-09-252-991A-25313
; Sequence 25313, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18

```

```

; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25313
; LENGTH: 582
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25313

Query Match 45.5%; Score 45; DB 4; Length 582;
Best Local Similarity 50.0%; Pred. No. 8.6;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 3 PGHERMGRGRTSSK 16
|||||I||:|I
Db 24 PGHRRAGRGRRGR 37

RESULT 11
US-09-252-991A-24639
; Sequence 24639, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMON.
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24639
; LENGTH: 940
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24639

Query Match 45.5%; Score 45; DB 4; Length 940;
Best Local Similarity 52.9%; Pred. No. 15;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 3 PGHERMGRGRTSSKLA 19
I:|||||I||:|I
Db 54 PGFRLFRGRSAAPELA 70

RESULT 12
US-09-252-991A-32512
; Sequence 32512, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMON.
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32512
; LENGTH: 266
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32512

Query Match 44.4%; Score 44; DB 4; Length 266;
Best Local Similarity 72.7%; Pred. No. 5.2;

```



```
Matches      8;  Conservative      0;  Mismatches      3;  Indels      0;  Gaps      0;

QY      3 PGHERMGRGT 13
DB      245 PGHRRGRGT 255

RESULT 13
US-09-252-991A-27955
; Sequence 27955, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27955
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27955

Query Match      44.4%; Score 44; DB 4; Length 341;
Best Local Similarity 53.8%; Pred. No. 6.9;
Matches      7;  Conservative      4;  Mismatches      2;  Indels      0;  Gaps      0;

QY      3 PGHERMGRGTSS 15
DB      87 PGADRLGGRTGA 99

RESULT 14
US-09-252-991A-16986
; Sequence 16986, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16986
; LENGTH: 417
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16986

Query Match      44.4%; Score 44; DB 4; Length 417;
Best Local Similarity 87.5%; Pred. No. 8.7;
Matches      7;  Conservative      1;  Mismatches      0;  Indels      0;  Gaps      0;

QY      3 PGHERMGR 10
DB      42 PGHORMGR 49

RESULT 15
US-09-252-991A-30947
; Sequence 30947, Application US/09252991A
; Patent No. 6551795
```

```
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30947
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30947
```

```
Query Match      44.4%; Score 44; DB 4; Length 530;
Best Local Similarity 72.7%; Pred. No. 12;
Matches      8;  Conservative      0;  Mismatches      3;  Indels      0;  Gaps      0;
```

```
QY      3 PGHERMGRGT 13
DB      70 PGHPRAGRGT 80
```

```
Search completed: October 1, 2003, 09:41:48
Job time : 19 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2003, 09:40:50 ; Search time 64 seconds

(without alignments)
46.969 Million cell updates/sec

Title: US-09-674-913A-1

Perfect score: 99

Sequence: 1 NVPCHRMGRGRTSSKELA 19

Scoring table:

BLOSUM62

Capop 10.0 , Gapext 0.5

Searched: 587654 seqs, 156212981 residues

Total number of hits satisfying chosen parameters: 587654

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	99	100.0	19	12 US-09-674-913A-1	Sequence 1, Appli
2	99	100.0	27	12 US-09-674-913A-2	Sequence 2, Appli
3	58	58.6	12	12 US-09-674-913A-4	Sequence 4, Appli
4	52	52.5	17	12 US-09-674-913A-3	Sequence 3, Appli
5	45	45.5	31	9 US-09-864-761-39237	Sequence 39237, A
6	45	45.5	70	11 US-09-764-891-4632	Sequence 4632, Ap
7	45	45.5	124	12 US-10-291-851-70	Sequence 70, Appl
8	43	43.4	480	15 US-10-156-761-11800	Sequence 11800, A
9	43	43.4	551	14 US-10-005-057A-2	Sequence 2, Appli
10	42	42.4	294	10 US-09-945-182-2	Sequence 2, Appli
11	42	42.4	353	15 US-10-092-263-8	Sequence 8, Appli
12	42	42.4	388	10 US-09-945-182-34	Sequence 34, Appli
13	42	42.4	411	10 US-09-945-182-28	Sequence 28, Appli
14	42	42.4	418	12 US-10-141-541-2	Sequence 2, Appli
15	42	42.4	450	15 US-10-188-246-12	Sequence 12, Appli

16	42	42.4	1285	12 US-10-141-541-4	Sequence 4, Appli
17	41	41.4	15	12 US-09-674-913A-5	Sequence 5, Appli
18	40.5	40.9	218	11 US-09-975-719-15	Sequence 15, Appli
19	40.5	40.9	498	15 US-10-156-761-14582	Sequence 14582, A
20	40.5	40.9	502	15 US-10-156-761-11689	Sequence 11689, A
21	40.5	40.9	510	15 US-10-156-761-11980	Sequence 11980, A
22	40	40.4	116	9 US-09-867-550-1898	Sequence 1898, Ap
23	40	40.4	283	15 US-10-270-875-52	Sequence 52, Appli
24	40	40.4	283	15 US-10-270-878-52	Sequence 52, Appli
25	40	40.4	283	15 US-10-270-786-52	Sequence 52, Appli
26	40	40.4	283	15 US-10-270-710-52	Sequence 52, Appli
27	40	40.4	283	15 US-10-270-859-52	Sequence 52, Appli
28	40	40.4	283	16 US-10-270-846-52	Sequence 52, Appli
29	40	40.4	322	12 US-10-238-075-330	Sequence 330, App
30	40	40.4	516	15 US-10-156-761-10628	Sequence 10628, A
31	39	39.4	120	11 US-09-974-879-607	Sequence 607, App
32	39	39.4	120	11 US-09-305-736-608	Sequence 608, App
33	39	39.4	202	16 US-10-090-170-63	Sequence 63, Appli
34	39	39.4	240	10 US-09-945-182-30	Sequence 30, Appli
35	39	39.4	247	10 US-09-738-626-5773	Sequence 5773, Ap
36	39	39.4	291	12 US-10-074-511-97	Sequence 97, Appli
37	39	39.4	368	15 US-10-156-761-14366	Sequence 14366, A
38	39	39.4	443	15 US-10-156-761-13401	Sequence 13401, A
39	39	39.4	538	15 US-10-013-803A-2	Sequence 2, Appli
40	39	39.4	1312	12 US-10-141-541-6	Sequence 6, Appli
41	39	39.4	1745	12 US-09-793-061-4	Sequence 4, Appli
42	39	39.4	2374	11 US-09-383-894-2	Sequence 2, Appli
43	39	39.4	2425	11 US-09-383-894-4	Sequence 4, Appli
44	38	38.4	53	9 US-09-864-761-43384	Sequence 43384, A
45	38	38.4	61	11 US-09-764-872-398	Sequence 398, App

ALIGNMENTS

RESULT 1

US-03-674-913A-1
; Sequence 1, Application US/09674913A
; Publication No. US20030171266A1
; GENERAL INFORMATION:
; APPLICANT: No. US20030171266A1sk Hydro ASA
; APPLICANT: Gaudernack, Gustav
; APPLICANT: Eriksen, Jon Amund
; APPLICANT: Moller, Mona
; TITLE OF INVENTION: Frameshift Mutants of Beta-Amyloid Precursor Protein and UB
; TITLE OF INVENTION: Their Use
; FILE REFERENCE: 001702.401600
; CURRENT APPLICATION NUMBER: US/09/674,913A
; CURRENT FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: PCT/NO99/00141
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: NO 19982098
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1
; LENGTH: 19
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-674-913A-1

Query Match 100.0%; Score 99; DB 12; Length 19;
Host Local Similarity 100.0%; Pred. No. 3.3e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NVPCHRMGRGRTSSKELA 19

Db 1 NVPCHRMGRGRTSSKELA 19

RESULT 2

US-09-674-913A-2
; Sequence 2, Application US/09674913A

```
; Publication No. US20030171266A1
; GENERAL INFORMATION:
; APPLICANT: Gaudernack, Gustav
; APPLICANT: Eriksen, Jon Amund
; TITLE OF INVENTION: Frameshift Mutants of Beta-Amyloid Precursor Protein and Ubiquitin
; FILE OF INVENTION: Their Use
; FILE REFERENCE: 001702.401600
; CURRENT FILING DATE: 2000-11-08
; PCT APPLICATION NUMBER: US/09/674,913A
; PCT FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: PCT/NO99/00141
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: NO 19982098
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 17
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-674-913A-3

Query Match          52.5%; Score 52; DB 12; Length 17;
Best Local Similarity 100.0%; Pred. No. 0.099;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NVPGHERMG 9
Db 9 NVPGHERMG 17
|||||

RESULT 5
US-09-864-761-39237
; Sequence 39237, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Pann, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL IN
; FILE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Agomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/532,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/235,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30

; Publication No. US20030171266A1
; GENERAL INFORMATION:
; APPLICANT: Gaudernack, Gustav
; APPLICANT: Eriksen, Jon Amund
; TITLE OF INVENTION: Frameshift Mutants of Beta-Amyloid Precursor Protein and Ubiquitin
; FILE OF INVENTION: Their Use
; FILE REFERENCE: 001702.401600
; CURRENT FILING DATE: 2000-11-08
; PCT APPLICATION NUMBER: US/09/674,913A
; PCT FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: PCT/NO99/00141
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: NO 19982098
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 27
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-674-913A-2

Query Match          100.0%; Score 99; DB 12; Length 27;
Best Local Similarity 100.0%; Pred. No. 4.7e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NVPGHERMGRTSSKELA 19
Db 9 NVPGHERMGRTSSKELA 27
|||||

RESULT 3
US-09-674-913A-4
; Sequence 4, Application US/09674913A
; Publication No. US20030171266A1
; GENERAL INFORMATION:
; APPLICANT: Gaudernack, Gustav
; APPLICANT: Eriksen, Jon Amund
; APPLICANT: Moller, Mona
; TITLE OF INVENTION: Frameshift Mutants of Beta-Amyloid Precursor Protein and Ubiquitin
; FILE OF INVENTION: Their Use
; FILE REFERENCE: 001702.401600
; CURRENT APPLICATION NUMBER: US/09/674,913A
; CURRENT FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: PCT/NO99/00141
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: NO 19982098
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-674-913A-4

Query Match          58.6%; Score 58; DB 12; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.0075;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 MGRGTSSKELA 19
Db 1 MGRGTSSKELA 12
|||||

RESULT 4
US-09-674-913A-3
; Sequence 3, Application US/09674913A
; Publication No. US20030171266A1
; GENERAL INFORMATION:
; APPLICANT: Gaudernack, Gustav
; APPLICANT: Eriksen, Jon Amund
; APPLICANT: Moller, Mona
; TITLE OF INVENTION: Frameshift Mutants of Beta-Amyloid Precursor Protein and Ubiquitin
; FILE OF INVENTION: Their Use
; FILE REFERENCE: 001702.401600
; CURRENT APPLICATION NUMBER: US/09/674,913A
; CURRENT FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: PCT/NO99/00141
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: NO 19982098
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 17
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-674-913A-3
```

```
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-25
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 39237
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006538.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4
US-09-864-761-39237

Query Match 45.5%; Score 45; DB 9; Length 31;
Best Local Similarity 50.0%; Pred. No. 2.5;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VPGHERMGGRITSSKELA 19
| | | | | : : :
DB 6 VPGHARPGHQSGSEAA 23

RESULT 6
US-09-764-891-4632
; Sequence 4632, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4632
; LENGTH: 70
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (85)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-4632

Query Match 45.5%; Score 45; DB 11; Length 70;
Best Local Similarity 56.2%; Pred. No. 5.8;
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 4 GHERMGGRITSSKELA 19
| | | | | : : :
DB 23 GVDVRVARGRENSKKLA 38

RESULT 7
US-10-291-851-70
; Sequence 70, Application US/10291851
; Publication No. US20030158384A1
; GENERAL INFORMATION:
; APPLICANT: Louie, Gordon V.
; APPLICANT: Gajiwala, Ketan S.
; APPLICANT: Buchanan, Sean Grant
```

```
; TITLE OF INVENTION: CRYSTALS AND STRUCTURES OF MEMBERS OF
; TITLE OF INVENTION: THE E. coli COMA AND YGBB PROTEIN FAMILIES (COMA)
; FILE REFERENCE: 52498200700
; CURRENT APPLICATION NUMBER: US/10/291,851
; CURRENT FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: 60/337,683
; PRIOR FILING DATE: 2001-11-09
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70
; LENGTH: 124
; TYPE: PRT
; ORGANISM: P.aeruginosa
US-10-291-851-70

Query Match 45.5%; Score 45; DB 12; Length 124;
Best Local Similarity 36.7%; Pred. No. 11;
Matches 11; Conservative 1; Mismatches 0; Indels 18; Gaps 1;

QY 2 VPGHERM-----GRGRT 13
| | | | | : | | | |
DB 60 LPGHERMARLSKLGTDLRVDYLRPGRGRT 89

RESULT 8
US-10-156-761-11800
; Sequence 11800, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 11800
; LENGTH: 480
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-11800

Query Match 43.4%; Score 43; DB 15; Length 480;
Best Local Similarity 80.3%; Pred. No. 92;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 VPGHERMGRG 11
| | | | | :
DB 96 VPGHERSGVG 105

RESULT 9
US-10-005-057A-2
; Sequence 2, Application US/10005057A
; Publication No. US20020170087A1
; GENERAL INFORMATION:
; APPLICANT: Tao, Yumin
; APPLICANT: Gordon-Kamm, William J.
; APPLICANT: Shen, RO
; APPLICANT: Lowe, Keith S.
; APPLICANT: Danilevskaya, Olga
; APPLICANT: Mahajan, Pramod
; APPLICANT: Rafalski, Antoni J.
; APPLICANT: Sakai, Hajime
```

APPLICANT: Klein, Ted M.
TITLE OF INVENTION: Transcriptional Regulator Nucleic Acids,
TITLE OF INVENTION: Polypeptides and Methods of Use Thereof
FILE REFERENCE: 1289
CURRENT APPLICATION NUMBER: US/00/005,057A
CURRENT FILING DATE: 2001-12-04
PRIOR APPLICATION NUMBER: 60/251,555
PRIOR FILING DATE: 2000-12-06
NUMBER OF SEQ ID NOS: 41
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 2
LENGTH: 551
TYPE: PRT
ORGANISM: Zea mays
US-10-005-057A-2

Query Match 43.4% Score 43; DB 14; Length 551;
Best Local Similarity 44.4% Pred. No. 1.1e+02;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 VPGHERMGRGRTSSKELA 19
DB 91 VEEHTAMGKGRSRKOWA 108

RESULT 10
US-09-945-182-2
Sequence 2, Application US/09945182
Patent No. US20020160494A1
GENERAL INFORMATION:
APPLICANT: Celeste, Anthony J.
Wozney, John
Rosen, Vicki A.
Wolfman, Neil
Thomsen, Gerald H.
Melton, Douglas A.

TITLE OF INVENTION: TENDON-INDUCING COMPOSITIONS
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENETICS INSTITUTE, INC.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/945,182
FILING DATE: 31-Aug-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/808,324
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: Lazar, Steven R.
REGISTRATION NUMBER: 32,618
REFERENCE/DOCKET NUMBER: 5202-D
TELEPHONE: 617 498-8260
TELEFAX: 617 876-5851

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 294 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-945-182-2

Query Match 42.4% Score 42; DB 10; Length 294;
Best Local Similarity 60.0% Pred. No. 80;
Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 GHERMGRGRTSSKEL 18
DB 184 GHGRHGRSRCSRKPL 198

RESULT 11
US-10-092-263-8
Sequence 8, Application US/10092263
Publication No. US2003006881A1
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: STEREOSELECTIVE REDUCTION OF SUBSTITUTED ACETOPHENONE
FILE REFERENCE: CT-2657NP
CURRENT APPLICATION NUMBER: US/10/092,263
CURRENT FILING DATE: 2002-06-04
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn version 3.0
SEQ ID NO 8
LENGTH: 353
TYPE: PRT
ORGANISM: Pichia methanolicus
US-10-092-263-8

Query Match 42.4% Score 42; DB 15; Length 353;
Best Local Similarity 44.4% Pred. No. 96;
Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 NVPGHERMGRGRTSSKEL 18
DB 115 NEPGHELVNRSLSRKHI 132

RESULT 12
US-09-945-182-34
Sequence 34, Application US/09945182
Patent No. US20020160494A1
GENERAL INFORMATION:
APPLICANT: Celeste, Anthony J.
Wozney, John
Rosen, Vicki A.
Wolfman, Neil
Thomsen, Gerald H.
Melton, Douglas A.
TITLE OF INVENTION: TENDON-INDUCING COMPOSITIONS
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENETICS INSTITUTE, INC.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/945,182
FILING DATE: 31-Aug-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/808,324
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Lazar, Steven R.
REGISTRATION NUMBER: 32,618
REFERENCE/DOCKET NUMBER: 5202-D
TELECOMMUNICATION INFORMATION:

```
;
; TELEPHONE: 617 498-8260
; TELEFAX: 617 876-5851
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 388 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-09-945-182-34
    Query Match          42.4%; Score 42; DB 10; Length 388;
    Best Local Similarity 60.0%; Pred. No. 1.1e+02;
    Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 GHERMGRGRTSSKEL 18
    | | | | |
Db 278 GHGRRGRSRCRKPL 232

RESULT 13
US-09-945-182-28
; Sequence 28, Application US/09945192
; Patent No. US20020160494A1
; GENERAL INFORMATION:
; APPLICANT: Celeste, Anthony J.
; Rosen, Vicki A.
; Wolman, Neil
; Thomsen, Gerald H.
; Melton, Douglas A.
; TITLE OF INVENTION: TENDON-INDUCING COMPOSITIONS
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENETICS INSTITUTE, INC.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/945.182
; FILING DATE: 31-Aug-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/808,324
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Lazar, Steven R.
; REGISTRATION NUMBER: 32,618
; REFERENCE/DOCKET NUMBER: 5202-D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617 498-8260
; TELEFAX: 617 876-5851
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 411 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 28:
US-09-945-182-28
    Query Match          42.4%; Score 42; DB 10; Length 411;
    Best Local Similarity 60.0%; Pred. No. 1.1e+02;
    Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 GHERMGRGRTSSKEL 18
```

```
Db 301 GHGRRGRSRCRKPL 315

RESULT 14
US-10-141-541-2
; Sequence 2, Application US/10141541
; Publication No. US20030167495A1
; GENERAL INFORMATION:
; APPLICANT: Pulst, Stefan M.
; TITLE OF INVENTION: SCA2 Knockout Animal and Methods of Use
; FILE REFERENCE: P-CF 5244
; CURRENT APPLICATION NUMBER: US/10/141,541
; PRIOR FILING DATE: 2002-05-07
; PRIOR APPLICATION NUMBER: US 60/289,231
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO: 2
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-141-541-2
    Query Match          42.4%; Score 42; DB 12; Length 418;
    Best Local Similarity 60.0%; Pred. No. 1.2e+02;
    Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 4 GHERMGRGRTSSKEL 18
    | | | | |
Db 54 GRPGLGRGNSSKGL 78

RESULT 15
US-10-188-246-12
; Sequence 12, Application US/10188246
; Publication No. US20030087274A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David W.,
; APPLICANT: Boldog, Ferenc,
; APPLICANT: Casman, Stacie,
; APPLICANT: Edinger, Shlomit,
; APPLICANT: Gerlach, Valerie,
; APPLICANT: Gorman, Linda,
; APPLICANT: Li, Li,
; APPLICANT: Malyankar, Uriel,
; APPLICANT: Patturajan, Meera,
; APPLICANT: Peyman, John,
; APPLICANT: Shenoy, Suresh,
; APPLICANT: Shimkets, Richard,
; APPLICANT: Vernet, Corine A. M.,
; APPLICANT: Voss, Edward
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND
; FILE REFERENCE: 21402-397B US
; CURRENT APPLICATION NUMBER: US/10/188,246
; CURRENT FILING DATE: 2002-07-02
; PRIOR APPLICATION NUMBER: 60/303046
; PRIOR FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 60/304502
; PRIOR FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 60/305011
; PRIOR FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: 60/306085
; PRIOR FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: 60/326981
; PRIOR FILING DATE: 2001-10-14
; PRIOR APPLICATION NUMBER: 60/360923
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: 60/363636
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/373063
; PRIOR FILING DATE: 2002-04-16
; NUMBER OF SEQ ID NOS: 60
```

```
; SOFTWARE: Custom
; SEQ ID NO 12
; LENGTH: 450
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-188-246-12

Query Match      42.4%   Score 42; DB 15; Length 450;
Best Local Similarity 60.0%   Pred. No. 1.2e+02;
Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY      4 GHERMGRGRTSSKEL 18
      ||| || ||| ||
Db      340 GHRRGRSRCSRKPL 354

Search completed: October 1, 2003, 09:50:21
Job time : 65 secs
```

GenCore version 5.1.6
 Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2003, 09:35:19 ; Search time 97 Seconds
 (without alignments)
 50.546 Million cell updates/sec

Title: US-09-674-913A-1
 Perfect score: 99
 Sequence: 1 NYPGHERMGRGRTSSKELA 19

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 330525

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriaph:*
- 17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	50.5	51.0	251	5 Q9V166	Q9V166 drosophila
2	50	50.5	72	16 Q9CIP3	Q9CIP3 lactococcus
3	46	46.5	342	16 Q97DL5	Q97DL5 clostridium
4	46	46.5	441	4 Q50149	Q50149 homo sapien
5	45.5	46.0	533	16 Q8CN90	Q8CN90 staphylococ
6	45	45.5	157	16 Q9HTU8	Q9HTU8 pseudomonas
7	45	45.5	331	10 Q9LH70	Q9LH70 arabidopsis
8	44.5	44.9	484	16 Q8XMY9	Q8XMY9 clostridium
9	44	44.4	2533	5 Q8IDH3	Q8IDH3 plasmodium
10	43.5	43.9	498	16 Q54116	Q54116 streptomyces
11	43	43.4	70	16 Q8E663	Q8E663 streptococ
12	43	43.4	70	16 Q8E018	Q8E018 streptococ
13	43	43.4	79	16 Q97RM2	Q97RM2 streptococ
14	43	43.4	79	16 Q8DQ62	Q8DQ62 streptococ
15	43	43.4	154	16 Q8PNB6	Q8PNB6 xanthomonas
16	43	43.4	244	4 Q8TF63	Q8TF63 homo sapien

17	43	43.4	424	3 Q74328	Q74328 schizosacch
18	43	43.4	441	11 Q99M11	Q99M11 mus musculu
19	42.5	42.9	491	16 Q97RA9	Q97RA9 streptococ
20	42.5	42.9	491	16 Q8DQ73	Q8DQ73 streptococ
21	42.5	42.9	1200	11 Q8CBV3	Q8CBV3 mus musculu
22	42	42.4	37	4 Q8XV6	Q8XV6 homo sapien
23	42	42.4	113	6 Q9NLS5	Q9NLS5 capreolus c
24	42	42.4	242	16 Q98608	Q98608 rhizobium l
25	42	42.4	294	6 Q9BDW9	Q9BDW9 macaca fasc
26	42	42.4	325	3 Q43042	Q43042 schizosacch
27	42	42.4	349	11 Q8R252	Q8R252 mus musculu
28	42	42.4	349	11 Q9D5G4	Q9D5G4 mus musculu
29	42	42.4	349	11 Q8Z883	Q8Z883 mus musculu
30	42	42.4	419	11 Q97421	Q97421 mus musculu
31	42	42.4	427	4 Q9ONY7	Q9ONY7 homo sapien
32	42	42.4	438	13 Q8UUP2	Q8UUP2 xenopus lae
33	42	42.4	441	11 Q921B6	Q921B6 mus musculu
34	42	42.4	447	6 Q9BDW8	Q9BDW8 cercopithe
35	42	42.4	481	17 Q97BP5	Q97BP5 thermoplasm
36	42	42.4	517	4 Q75530	Q75530 homo sapien
37	42	42.4	535	11 Q97452	Q97452 mus musculu
38	42	42.4	586	16 Q9A8S0	Q9A8S0 caulobacter
39	42	42.4	821	6 Q19060	Q19060 saguinus oe
40	42	42.4	961	16 Q8YQV0	Q8YQV0 anabaena sp
41	42	42.4	962	16 Q9JYD2	Q9JYD2 neisseria m
42	42	42.4	962	16 Q9J7B5	Q9J7B5 neisseria m
43	42	42.4	1084	10 Q8SB87	Q8SB87 oryza sativ
44	42	42.4	1285	11 Q70305	Q70305 mus musculu
45	41.5	41.9	236	16 Q9KZM4	Q9KZM4 streptomyces

ALIGNMENTS

RESULT: 1

Q9V166 PRELIMINARY: PRT: 251 AA.

AC Q9V166: Q95SF6:

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DI 01-OCT-2002 (TREMBLrel. 22, Last sequence update)

DI 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE CG31248 protein (GH2863p).

GN CG31248 OR CG2640.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Ephydroidea; Ephydroidea; Ephydroidea;

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Berkeley;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amaralides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chai L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne B.D.,

Wan K.H., Doyle C., Baxter F.G., Helt J., Nelson C.R., Pfeiffer G.L.,

Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Balowin D.,

Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

Reeson K.Y., Henos P.V., Berman B.P., Bhandari D., Bolshakov S.,

Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,

Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

Codson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

Durlin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

Glock K., Gong F., Gorrelli J.H., Gu Z., Guan P., Harris M.,

Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwa C.,

Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

LSJ :PSHERUGIGRSLNSQPTIELTKEIA 177

RESULT 3
QCICP3 PRELIMINARY; PRT: 72 AA.
ID QCIP3
AC QCIP3;
CT C1-JUN-2001 (TREMBLrel.. 17, Created)
DT 01-JUN-2001 (TREMBLrel.. 17, Last sequence update)
DI 01-MAR-2002 (TREMBLrel.. 20, Last annotation update)
DE Hypothetical protein ydbC.
GN YDC OK L:0313
GS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OS Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OC NCBI_taxid=1360;
OX [1]
RP SEQUENCE FROM N.A.
RN STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Manger S., Jaillon O., Malarme K.,
RC Weissenbach J., Ehrlich S.D., Sorokin A.;
RD "Lactis ssp. lactis IL1403."
RE Lactis ssp. lactis IL1403.";
RL Genome Res. 11:731-753(2001).
RS EMBL: AE006268; AAC04411.1; .
RW Hypothetical protein; Complete proteome
KW Hypothetical protein; Complete proteome
SQ SEQUENCE 72 AA: 8405 MW; CS041FC4543B8410 CRC64;

Query Match 50.5%; Score 50; DB 16; Length 72;
Best Local Similarity 60.0%; Pred.No. 0.88;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps

QY 3 PGHFMGRGHTSSKE 17
| | ||||| | | | |
DB 45 PDHEKMGKGITLSEE 60

RESULT 3
QC7DL5 PRELIMINARY; PRT: 342 AA.
ID Q7DLS
AC Q7DLS;
CT 01-OCT-2001 (TREMBLrel.. 18, Created)
DT 01-OCT-2001 (TREMBLrel.. 18, Last sequence update)
DI 01-NAR-2003 (TREMBLrel.. 23, Last annotation update)
DE Homolog of cell division GTPase FtsZ, diverged.
GN CA3345,
GS Clostridium acetobutylicum.
OS Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCB-taxid=1488;
OY [1]
RP SEQUENCE FROM N.A.
RN STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RB Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J.W., Wolf Y.I.,
RC Ratsov R.L., Sabatche F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RD Bennett G.N., Koehn E.V., Smith D.R.;
RE "Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum";
RF J. Bacteriol. 183:4823-4838(2001).
RG FMAL: AE007843; AAA81388.1;
RH InterPro: IPR000217; Tubulin.
RI PROSITE: PS00227; TUBULIN: 1.
RK Cell division; Complete proteome.
SQ SEQUENCE 342 AA: 37499 MW; D2288911AA0485F00 CRC64;

```

Query Match      46.5%; Score 46; DB 16; Length 342;
Best Local Similarity 56.3%; Pred. No. 21;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

CY      3  PGERHMGRTSSKEL 18
      || : || || || :

```

```

Db 57 PGTDGSGDRDRKSKEM 72

RESULT 4
ID00149 PRELIMINARY: PRT: 441 AA.
AC 000149;
DT 01-JUL-1997 (TRENBLrel. 04, Created)
DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Embryonic ECTODERM development protein homolog.
GN EED.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RA Peytavi R., Hong S.S., Gay B., Dupuy D'Angeac A., Seliq L., Benichou S.,
RA Benichou S., Benarous K., Boulanger P.;
RT "Heed, the product of the human homolog of the murine eed gene, binds
RT to the matrix protein of HIV-1.";
RL J. Biol. Chem. 0:0-C(1998).
[2]
RN SEQUENCE OF 1-400 FROM N.A.
RA Peytavi R., Hong S.S., Gay B., d'Angeac A.D., Seliq L., Benichou S.,
RA Benarous K., Boulanger P.;
RT "HEED, the product of the human homolog of the murine eed gene, binds
RT to the matrix protein of HIV-1.";
RL J. Biol. Chem. 274:1635-1645(1999).
DR EMBL: U90651; AAD08714.1;
DR EMBL: AF099032; AAD08815.1;
DR InterPro: IPR000408; Reg_chnr_Condens.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 5.
DR SMART: SM00320; WD40; 3.
DR PROSITE: PS00626; RC1.2; 1.
DR PROSITE: PS00678; WD_REPEATS_1; 1.
DR PROSITE: PS00882; WD_REPEATS_2; 2.
DR PROSITE: PS02094; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 441 AA; 50155 MW; 2EA45A5BEEEA56B0 CRC64;

Query Match 46.5%; Score 46; DB 4; Length 441;
Best Local Similarity 47.1%; Pred. No. 27;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 1 NVPGHERMGRGRTSSKE 17
| | | | | | | | | | | | | | | | |
Db 61 NPGKSGKSGKWSKK 77

RESULT 5
QBCN90 PRELIMINARY: PRT: 533 AA.
AC QBCN90;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE L-lactate permease lctP-like protein.
GN SE1945.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1292;
[1]
RN SEQUENCE FROM N.A.
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
RA Chen Z., Wen Y.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL: AE016750; AAO05586.1;

KW Complete proteome.
SQ SEQUENCE 533 AA; 56639 MW; BB890AFA98D7487C CRC64;

Query Match 45.0%; Score 45.5; DB 16; Length 533;
Best Local Similarity 55.0%; Pred. No. 40;
Matches 11; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

Qy 1 NV7GH-ERMGRGRTSSKELA 19
| | | | | | | | | | | | | | | | |
Db 182 NLPGRVEMGVSTSTLILA 201

RESULT 6
Q9HTU8 PRELIMINARY: PRT: 157 AA.
AC Q9HTU8;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Hypothetical protein PA5246.
GN PA5246.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
[1]
RN SEQUENCE FROM N.A.
RA STRAIN-ATCC 15692 / PA01;
RA MEDLINE=2043737; PubMed=10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warrenner F.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Carber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reiter J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL: AE004937; AAG08631.1;
DR InterPro: IPR003736; DUF157.
DR InterPro: IPR006683; Thioestr_supf.
DR Pfam: PF03061; 4HBT; 1.
DR TIGRfams: TIGR00369; unchar_dom_1; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 157 AA; 17005 MW; 1002F16B858253A3 CRC64;

Query Match 45.5%; Score 45; DB 16; Length 157;
Best Local Similarity 36.7%; Pred. No. 14;
Matches 11; Conservative 1; Mismatches 0; Indels 18; Gaps 1;

Qy 2 VPGHERM-----GRGRT 13
| | | | | | | | | | | | |
Db 85 LPGHERMARLSKLTIDLRVDYLRPGRGRT 118

RESULT 7
Q9LH70 PRELIMINARY: PRT: 331 AA.
AC Q9LH70;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE GB|AAD2396.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
[1]
RN SEQUENCE FROM N.A.
RA STRAIN-Columbia.
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.

```

```

RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Colombia;
RX  MEDLINE=20363099; PubMed=0907853;
RA  Nakamura Y.;
RT  "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RI  Sequence features of the regions of 4,251,695 bp covered by nine-y P1.
RT  TAC and BAC clones.";
RL  DNA Res. 7:217-221(2000).
DR  EMBL; AP002061; BAB02651.1; -.
DR  InterPro; IPR004253; DUF231.
DR  Pfam; PF03005; DUF231.1
SQ  SEQUENCE 331 AA; 37697 MW; 165BE4B9EFFEB9B8 CRC64;

Query Match 45.5%; Score 45; DB 10; Length 331;
Best Local Similarity 56.2%; Pred. No. 30;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY  3 PGERMGRGRTSSKEL 18
DB  253 PSHYKSSVGRSGREL 268

RESULT 8
OBXMY9
ID  OBXMY9 PRELIMINARY; PRT; 484 AA.
AC  OBXMY9
DT  01-MAR-2002 (TRENBLrel. 20, Created)
DT  01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT  01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE  Lysine decarboxylase.
GN  DCLY OR CPE0549.
OS  Clostridium perfringens.
OC  Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC  Clostridium.
OX  NCBI_TaxID=1502;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=13 / Type A;
RX  PubMed=11792842;
RA  Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA  Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT  "Complete genome sequence of Clostridium perfringens, an anaerobic
RT  flesh-eater.";
RL  Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR  EMBL; AP003187; BAB0255.1; -.
DR  InterPro; IPR000310; Decarboxylase.
DR  Pfam; PF01276; OKR_DC_1; 1.
DR  Pfam; PF03711; OKR_DC_1-C; 1.
DR  PROSITE; PS00703; OKR_DC_1; 1.
KW  Complete proteome.
SQ  SEQUENCE 484 AA; 53971 MW; E33933E34C92EFAC CRC64;

Query Match 44.9%; Score 44.5; DB 16; Length 484;
Best Local Similarity 52.9%; Pred. No. 54;
Matches 9; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

QY  1 NYPGHERMGRGRTSSKE 17
DB  27 DVPGH-KQGRGNTELRE 42

RESULT 9
OBIDH3
ID  OBIDH3 PRELIMINARY; PRT; 2533 AA.
AC  OBIDH3;
DT  01-MAR-2003 (TRENBLrel. 23, Created)
DT  01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT  01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE  Hypothetical protein, conserved.
GN  PF13_0273.
OS  Plasmodium falciparum (isolate 3D7).
OC  Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

```

```

OX  NCBI_TaxID=36329;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Harris B., Lennard N., Clark L., Line A., Barton A., Corton C.,
RA  Herriman M., Paig A., Hall N., Atkin R., Chillingworth C., Doggett J.,
RA  Crimond D., Sanders M., Hayes R., Hall S., Quail M., Harrell B.,
RA  Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AL844509; CAD52648.1; -.
KW  Hypothetical protein.
SQ  SEQUENCE 2533 AA; 304302 MW; 42BB01D2645D26BD CRC64;

Query Match 44.4%; Score 44; DB 5; Length 2533;
Best Local Similarity 43.8%; Pred. No. 3.7e+02;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY  1 NYPGHERMGRGRTSSK 16
DB  408 NIPNHDKKKKRSSTK 423

RESULT 10
OS4116
ID  OS4116 PRELIMINARY; PRT; 498 AA.
AC  OS4116;
DT  01-JUN-1998 (TRENBLrel. 06, Created)
DT  01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DT  01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE  SC10A5.25c protein.
GN  SC05920 OR SC10A5.25c.
OS  Streptomyces coelicolor.
OC  Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC  Streptomycineae; Streptomycetaceae; Streptomyces.
OX  NCBI_TaxID=1902;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=A3(2);
RA  Murphy L., Harris D.;
RL  Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=A3(2);
RA  Parkhill J., Barrell B.G., Rajandream M.A.;
RI  Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN  [3]
RP  SEQUENCE FROM N.A.
RC  STRAIN=A3(2);
RX  MEDLINE=9700351; PubMed=8843436;
RA  Redenbach M., Kieser H.M., Denepaite D., Eichner A., Cullum J.,
RA  Kinashi H., Hopwood D.A.;
RT  "A set of ordered cosmids and a detailed genetic and physical map for
RT  the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL  Mol. Microbiol. 21:77-96(1996).
RN  [4]
RP  SEQUENCE FROM N.A.
RC  STRAIN=A3(2) / M145;
RX  MEDLINE=21596410; PubMed=12000953;
RA  Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA  Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA  Harper D., Batsman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA  Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA  Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA  Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA  Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA  Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA  Hopwood D.A.;
RI  "Complete genome sequence of the model actinomycete Streptomyces
RI  coelicolor A3(2).";
RJ  Nature 417:141-147(2002).
DR  EMBL; AL939125; CAAL6457.1; -.
DR  HSSP; Q58083; 1HV8.
DR  InterPro; IPR001410; DEAD.
DR  InterPro; IPR001650; Helicase_C.
DR  InterPro; IPR001254; Ser_protease_Try.

```

```

DR Pfam: PF00270; DEAD: 1.
DR Pfam: PF00271; helicase_C; 1.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELIC; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW ATP-binding; Helicase; Complete proteome.
SQ SEQUENCE 498 AA; 52678 MW; 79F02BC9D5C96D10 CRC64;

Query Match 43.9%; Score 43.5; DB 16; Length 498;
Best Local Similarity 55.0%; Pred. No. 81;
Matches 11; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

OY 1 NVPGERMGRGT-SSXELA 19
   : : : : : : : : : : : :
DB 103 SLAGKIMGRGTSGKTLA 122

RESULT 11
ID OBE663 PRELIMINARY; PRT; 70 AA.
AC OBE663;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Hypothetical protein.
GN GB50764.
OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=2.6495;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NEW316 / Serotype I-I;
RX MEDLINE=22242508; PubMed=12354221;
RA Glaser P., Rusnlok C., Buchrieser C., Chevalier F., Frangeul L.,
RA Msadek T., Zouine M., Couve E., Lalioui L., Poyart C., Trieu-Cuot P.,
RA Kunst F.;
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT invasive neonatal disease.";
RL Mol. Microbiol. 45:1499-1513(2002).
DR EMBL: AL766846; CAD46408.1; -.
DR SAGALIST; GDS0764; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 70 AA; 8106 MW; 693AF5432945686F CRC64;

Query Match 43.4%; Score 43; DB 16; Length 70;
Best Local Similarity 53.3%; Pred. No. 13;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 3 PGHERMGRGRTSSKE 17
   : : : : : : : : : : : :
DB 45 PDHTKMGKGITLSNE 59

RESULT 12
ID Q8E018 PRELIMINARY; PRT; 70 AA.
AC Q8E018;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Conserved hypothetical protein.
GN SAG0743.
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=22222988; PubMed=12200547;
RA Tettelin H., Massignani V., Cieslewicz M.J., Eisen J.A., Peterson S.,
RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,

```

```

RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S.,
RA Carly H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Zacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Malone D.,
RA Rinaldo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
DR EMBL: AE014225; AAM99630.1; -.
DR TIGR: SAG0743; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 70 AA; 8106 MW; 693AF5432945686F CRC64;

Query Match 43.4%; Score 43; DB 16; Length 70;
Best Local Similarity 53.3%; Pred. No. 13;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 3 PGHERMGRGRTSSKE 17
   : : : : : : : : : : : :
DB 45 PDHTKMGKGITLSNE 59

RESULT 13
ID Q97RM2 PRELIMINARY; PRT; 79 AA.
AC Q97RM2;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical protein SP0782.
GN SP0782.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Haft D.H., Dodson R.J.,
RA Peterson S., Heidelberg J., DeBoy R.T., Nelson W.C., Peterson J.D.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson M.R., Radune D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum I.V., Angiuoli S., Dickinson I., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith R.O., Venter J.C.,
RA Dougherty B.A., Morrison B.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:458-506(2001).
DR EMBL: AF007385; AAK74919.1; -.
DR TIGR: SP0782; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 79 AA; 9136 MW; 7CEAE8957946144B CRC64;

Query Match 43.4%; Score 43; DB 16; Length 79;
Best Local Similarity 53.3%; Pred. No. 14;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 3 PGHERMGRGRTSSKE 17
   : : : : : : : : : : : :
DB 53 PDHTKMGKGITLSNE 67

RESULT 14
ID Q8DQG2 PRELIMINARY; PRT; 79 AA.
AC Q8DQG2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

```

DE Conserved hypothetical protein.

GN SP0690.

OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).

OC Bacteria: Firmicutes: Lactobacillales: Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=171101;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21429245; PubMed=11544234;

RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,

RA DeHoff B.S., Estrom S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,

RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Laquer R.E.,

RA LeBlanc D.J., Lee L.N., Letkowitz E.J., Lu J., Matsushima P.,

RA McAhren S.M., McHenry M., McLeaster K., Mundy C.W., Niclas I.I.,

RA Norris F.H., O'Gata M., Peery R.B., Robertson G.L., Rockey P.,

RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,

RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostock P.R. Jr., Skatrud P.L.,

RA Glass J.I.;

RT *Genome of the bacterium Streptococcus pneumoniae strain R6.*;

RL J. Bacteriol. 183:5709-5717(2001).

DR EMBL: AE008446; AAK99494.1; ..

KW Hypothetical protein: Complete proteome.

SQ SEQUENCE 79 AA: 9136 MW: 7CEAE8957946144B CRC64;

Query Match 43.4%; Score 43; DB 16; Length 79;

Best Local Similarity 53.3%; Pred. No. 14;

Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 PGERMGGRGRTSSKE 17

DB 53 PDHTRMGKGITLSNE 67

RESULT 15

Q8PNB6

ID Q8PNB6 PRELIMINARY; PRT: 154 AA.

AC Q8PNB6;

DT 01-OCT-2002 (TREMBLrel. 22, Created)

DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)

DE 01-OCT-2002 (TREMBLrel. 22, Last annotation update)

DE Hypothetical protein XAC1157.

GN XAC1157.

OS Xanthomonas axonopodis (pv. citri).

OC Bacteria: Proteobacteria: Gammaproteobacteria: Xanthomonadales;

OC Xanthomonadaceae; Xanthomonas.

OX NCBI_TaxID=92829;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=306 / ATCC 13902 / XV 101;

RX MEDLINE=22022145; PubMed=12024217;

RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furian L.R.,

RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,

RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.F.A.,

RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.F.,

RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,

RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.I.,

RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,

RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,

RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,

RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,

RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,

RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,

RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira F.C., Tezza R.I.D.,

RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,

RA Setubal J.C., Kitajima J.P.;

RT *Comparison of the genomes of two Xanthomonas pathogens with differing

host specificities.*;

RL Nature 417:459-463(2002).

DR EMBL: AE011745; AAM36029.1; ..

KW Hypothetical protein: Complete proteome.

SQ SEQUENCE 154 AA: 17260 MW: FC8DDF515D2D7715 CRC64;

Query Match

43.4%; Score 43; DB 16; Length 154;

Best Local Similarity 61.5%; Pred. No. 29;

Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 PGERMGGRGRTSS 15

DB 111 : 111 11

DB 7 FGRALRRGRRSS 19

Search completed: October 1, 2003, 09:40:41

Job time : 102 secs

GenCore version 5.1.6
Copy: right (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2003, 09:34:45 : Search time 22 seconds
(without alignments)
40.614 Million cell updates/sec

Title: US-09-674-913a-1

Perfect score: 99
Sequence: 1 NYPGHERMGRGRTSSKELA 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	ID	Description
1	44.5	44.9	1 RCA_ANASP	P58555 anabaena sp
2	43	43.4	1 GDF7_MOUSE	P43029 mus musculus
3	42.5	42.9	1 RCA_ANASC	Q06721 anabaena sp
4	42	42.4	1 RECA_RHOCA	P42447 rhodobacter
5	42	42.4	1 RECA_PARDE	P95469 paracoccus
6	41	41.4	1 RSI3_MYCHO	P45813 mycobacteri
7	41	41.4	1 RECA_ACHLA	P29225 acetoelplasm
8	41	41.4	1 RECA_BACAN	Q9ap22 bacillus an
9	40	40.4	1 IGF2_POVIN	P07456 bos taurus
10	40	40.4	1 IGF2_SHEEP	P10764 ovis aries
11	40	40.4	1 RECA_RHOSH	P32725 rhodobacter
12	40	40.4	1 CBIA_SALT_Y	P29946 salmonella
13	40	40.4	1 REB1_SCHPO	Q9p6h9 schizosacch
14	40	40.4	1 ROM_HUMAN	P52272 homo sapien
15	39.5	39.9	1 POLX_TOBAC	P10978 nicotiana t
16	39	39.4	1 CAS1_PIG	P39035 sus scrofa
17	39	39.4	1 RNC_CORGL	Q8nnv6 corynebacte
18	39	39.4	1 CYSM_BACSU	Q34476 bacillus su
19	39	39.4	1 MID_PTCCR	P42754 petroselinu
20	39	39.4	1 RECA_PORGI	P77925 porphyromon
21	39	39.4	1 RECA_CHLTE	O52393 chlorobium
22	39	39.4	1 YARK_ECOLI	P75691 escherichia
23	39	39.4	1 D12_GREAL	O81931 crepis aipi
24	39	39.4	1 PRIM_PSEAE	Q915w0 pseudomonas
25	39	39.4	1 XLNR_ASPNG	C42804 aspergillus
26	39	39.4	1 RPOD_SINAL	Q9thv5 sinapis alb
27	39	39.4	1 CCAS_RAT	O54898 rattus norv
28	38	38.4	1 Y4WH_RH1SN	P55686 rhizobium s
29	38	38.4	1 IGF2_PIG	P23695 sus scrofa
30	38	38.4	1 RS4_PSEAE	O52759 pseudomonas
31	38	38.4	1 C9U7_HUMAN	Q9btj7 homo sapien
32	38	38.4	1 FLAG_BORBU	O51715 borrelia bu
33	38	38.4	1 PANB_CORGL	Q9x712 corynebacte

ALIGNMENTS

RESULT 1			
PCA_ANASP			
ID	RCA_ANASP	STANDARD:	PRT: 414 AA.
AC	P58555:		
DT	28-FEB-2003	(Rel. 41, Created)	
LT	28-FEB-2003	(Rel. 41, Last sequence update)	
LT	28-FEB-2003	(Rel. 41, Last annotation update)	
DE	Ribulose biphosphate carboxylase/oxygenase activase (RuBisCO		
DE	activase) (RA)		
GN	RCA OR ALR1533.		
OS	Anabaena sp. (strain PCC 7120).		
OC	Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.		
OX	NCBI_TaxID=103690;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE-21595285; Pubmed=11759840;		
RA	Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,		
RA	Katanahe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,		
RA	Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,		
RA	Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,		
RA	Yasuda M., Tabata S.;		
RT	*Complete genomic sequence of the filamentous nitrogen-fixing		
RT	cyanobacterium Anabaena sp. strain PCC 7120.*;		
RL	DNA Res. 8:205-213(2001).		
CC	-1- FUNCTION: ACTIVATION OF RUBISCO (RUBISCO-1.5-BISPHOSPHATE		
CC	CARBOXYLASE/OXYGENASE; EC 4.1.1.39) INVOLVES THE ATP-DEPENDENT		
CC	CARBOXYLATION OF THE EPSILON-AMINO GROUP OF LYSINE LEADING TO A		
CC	CARBAMATE STRUCTURE (BY SIMILARITY).		
CC	-1- SIMILARITY: BELONGS TO THE RUBISCO ACTIVASE FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercia		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; AP003586; BAB77899.1; -		
DR	PIR; AG1997; AG1997.		
DR	InterPro; IPR003959; AAA_Attrase_cent.		
DR	Pfam; PF00004; AAA; 1.		
DR	ATP-binding; Complete proteome.		
FT	NE_BIND 37 44 ATP (POTENTIAL)		
ST	SEQUENCE 414 AA: 46668 MW: E8359F10DBCS58DA CRC64:		
Query Match 44.9% Score 44.5; DB 1; Length 414;			
Best Local Similarity 39.3% Pred. No. 5.3;			
Matches 11; Conservative 2; Mismatches 6; Indels 9; Gaps 1;			
Qy	1 NYPG-----HERMGRGRTSSKELA 19		
	: :		
Db	23 NIPGVRVPLILGTHGRKGRKTFQCELA 50		
	: :		

Q8uj05 agrobacteri
P42751 arabisdopsis
P16971 bacillus su
P33542 aquifex pyr
Q9kaa7 bacillus ha
C52394 chloroflexu
P42734 arabisdopsis
Q07103 pyrospora
Q8u3a0 pyrococcus
Q01587 cucumis sat
O87392 rhizobium m
P12457 euglena gra

```

GDF7_MOUSE ID GDF7_MOUSE STANDARD: PRT: 151 AA.
AC P43029;
DT 01-NOV-1995 (Rel. 32, Created)
DI 16-OCT-2001 (Rel. 40, Last sequence update)
DE Growth/differentiation factor 7 precursor (GDF-7) (Fragment).
GN GDF7 OR GDF-7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BALB/c; TISSUE=Liver;
RC MEDLINE=94195427; PubMed=8145850;
RA Storm E.E., Huynh T.V., Copeland N.G., Jenkins N.A., Kingsley D.M.,
RA Lee S.-J.;
RT *Limb alterations in brachypodism mice due to mutations in a new
RT member of the TGF beta-superfamily.*;
RL Nature 368:639-643(1994).
CC -!- SUBUNIT: Homodimer; disulfide-linked (By similarity).
CC -!- SIMILARITY: Belongs to the TGF-beta family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U08339; AAA18780.1;
DR PIR: S43296; S43296.
DR HSP: P12643; 38MP.
DR MGD; MGI:95690; Gdf7.
DR InterPro: IPR002400; GF_cysknot.
DR InterPro: IPR001839; TGFb.
DR Pfam: PFC0019; TGF-beta; 1.
DR PRINTS; PR00438; GFCYSKNOT.
DR ProDom; PD000357; TGPB; 1.
DR SMART; SM00204; TGFb; 1.
DR PROSITE; PS00250; TGF_beta_1; 1.
KW Growth factor; Cytokine; Glycoprotein.
FT NON_TER 1 1
FT PROPEP <1 5 POTENTIAL.
FT CHAIN 6 151 GROWTH/DIFFERENTIATION FACTOR 7.
FT DISULFID 50 116 BY SIMILARITY.
FT DISULFID 79 148 BY SIMILARITY.
FT DISULFID 83 150 BY SIMILARITY.
FT DISULFID 115 115 INTERCHAIN (BY SIMILARITY).
FT DOMAIN 1 5 POLY-ARG.
FT DOMAIN 16 41 POLY-GLY.
FT SEQUENCE 151 AA; 15697 MW; 0E496ACB5827759 CRC64;

Query Match 43.4%; Score 43; DB 1; Length 151;
Best Local Similarity 60.0%; Pred. No. 3.2;
Matches 9; Conservative 0; Mismatches 6; Indels 2; Gaps 0;

QY 4 GHRMGRGRTSSKEL 18
DB 41 GHGRRGRSRSKSL 55

RESULT 3
RCA_ANASC ID RCA_ANASC STANDARD: PRT: 415 AA.
AC Q06721;
DT 16-OCT-2001 (Rel. 40, Created)
DI 16-OCT-2001 (Rel. 40, Last sequence update)
DE Ribulose biphosphate carboxylase/oxygenase activase (RuBisCo
DE activase) (RA).

```

```

RCA_Anabaena sp. (strain CA / ATCC 33047).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
OX NCBI_TaxID=52271;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93222475; PubMed=8467074;
RA L.L.A., Gibson J.L., Tabita F.R.;
RT "The Rubisco activase (rca) gene is located downstream from rbcS in
RT Anabaena sp. strain CA and is detected in other Anabaena/Nostoc
RT strains.*";
RL Plant Mol. Biol. 21:753-764(1993).
CC -!- FUNCTION: ACTIVATION OF RUBISCO (RUBULOSE-1,5-BISPHOSPHATE
CC CARBOXYLASE/OXYGENASE; EC 4.1.1.39) INVOLVES THE ATP-DEPENDENT
CC CARBOXYLATION OF THE EPSILON-AMINO GROUP OF LYSINE LEADING TO A
CC CARBAMATE STRUCTURE (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE RUBISCO ACTIVASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X67942; CAA48129.1;
DR InterPro: IPR003959; AAA_Atpase_centri.
DR Pfam; PF00004; AAA; 1.
KW ATP-binding. 37 44
FT NP_BINS 37 44 ATP (POTENTIAL).
FT SEQUENCE 415 AA; 46594 MW; 1CF71296D94A892D CRC64;

Query Match 42.9%; Score 42.5; DB 1; Length 415;
Best Local Similarity 39.3%; Pred. No. 12;
Matches 11; Conservative 2; Mismatches 6; Indels 9; Gaps 1;

QY 1 NVPS-----HERMGRGRTSSKELA 19
DB 23 NLPGRVPLIIGIHGRKGKGFQCEIA 50

RESULT 4
RCA_RHOCA ID RCA_RHOCA STANDARD: PRT: 355 AA.
AC P4247;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 26-FEB-2003 (Rel. 41, Last annotation update)
DE Rca protein (Recombinase A).
GN RCA.
CS Rhodobacter capsulatus (Rhodospseudomonas capsulata).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
OC Rhodobacteraceae; Rhodobacter.
OX NCBI_TaxID=10611;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=J50;
RX MEDLINE=95331467; PubMed=7607398;
RA Fernandez de Henestrosa A.R., Rivera E., Barbe J.;
RT "Non-reciprocal regulation of Rhodobacter capsulatus and Rhodobacter
RT sphaeroides rca genes expression.*";
RL FEMS Microbiol. Lett. 129:175-181(1995).
CC -!- FUNCTION: CAN CATALYZE THE HYDROLYSIS OF ATP IN THE PRESENCE OF
CC SINGLE-STRANDED DNA, THE ATP-DEPENDENT HYBRIDIZATION OF
CC DNA BY DUPLEX DNA, AND THE ATP-DEPENDENT HYBRIDIZATION OF
CC HOMOLOGOUS SINGLE-STRANDED DNAs. IT INTERACTS WITH LEXA CAUSING
CC ITS ACTIVATION AND LEADING TO ITS AUTOCATALYTIC CLEAVAGE.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the rca family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```


RESULT 7

RECA_ACHLA ID RECA_ACHLA STANDARD: PRI: 331 AA.
 AC P29225;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE RECA protein (Recombinase A).
 GN RECA.
 OS Acholeplasma laidlawii.
 OC Bacteria; Firmicutes; Mollicutes; Acholeplasmatales;
 OC Acholeplasmataceae; Acholeplasma.
 OX NCBI_TaxID=2148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL, K2, and 8195;
 RX MEDLINE=92121115; PubMed=1732213;
 RA Dybvig K., Woodard A.;
 RT "Cloning and DNA sequence of a mycoplasmal recA gene.";
 RL J. Bacteriol. 174:778-784(1992).
 CC -|- FUNCTION: CAN CATALYZE THE HYDROLYSIS OF ATP IN THE PRESENCE OF SINGLE-STRANDED DNA, THE ATP-DEPENDENT UPTAKE OF SINGLE-STRANDED DNA BY DUPLEX DNA, AND THE ATP-DEPENDENT HYBRIDIZATION OF HOMOLOGOUS SINGLE-STRANDED DNAs. IT INTERACTS WITH LEXA CAUSING ITS ACTIVATION AND LEADING TO ITS AUTOCATALYTIC CLEAVAGE.
 CC -|- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -|- SIMILARITY: Belongs to the recA family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M81465; AAB59011.1; -;
 DR PIR: A42602; A42602.
 DR HSSP: P26345; IG18.
 DR HAMAP: MF_00268; -; 1.
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR001553; RecA.
 DR Pfam: PF00154; RecA; 1.
 DR PRINTS: PR00142; RECA.
 DR ProDom: PD000229; RecA; 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00321; RECA_1; 1.
 DR PROSITE: PS50162; RECA_2; 1.
 DR PROSITE: PS50163; RECA_3; 1.
 DR DNA damage; DNA recombination; SOS response; ATP-binding; DNA-binding.
 FT NP_BIND 66 73
 FT VARIANT 295 331 MISSING (IN STRAIN 8195).
 SQ SEQUENCE 331 AA; 35516 MW; 60F3EA59CED83FF2 CRC64;

Query Match 41.4%; Score 4i: DB 1; Length 331;
 Best Local Similarity 42.9%; Pred. No. 16;
 Matches 6; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 4 GHERMGRGRTSSKE 17

| : : : : : : : |

Db 294 GEEKLGGRDNNAQ 307

RESULT 8

RECA_BACAN ID RECA_BACAN STANDARD: PRI: 343 AA.
 AC Q9AP22;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE RecA protein (Recombinase A).
 GN RECA.
 OS Bacillus anthracis.

CC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 CX NCBI_TaxID=1392;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sterne;
 RA Ko M., Kim J.C., Park C.;
 RT "Cloning of recA gene from Bacillus anthracis.";
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 CC -|- FUNCTION: CAN CATALYZE THE HYDROLYSIS OF ATP IN THE PRESENCE OF SINGLE-STRANDED DNA, THE ATP-DEPENDENT UPTAKE OF SINGLE-STRANDED DNA BY DUPLEX DNA, AND THE ATP-DEPENDENT HYBRIDIZATION OF HOMOLOGOUS SINGLE-STRANDED DNAs. IT INTERACTS WITH LEXA CAUSING ITS ACTIVATION AND LEADING TO ITS AUTOCATALYTIC CLEAVAGE.
 CC -|- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -|- SIMILARITY: Belongs to the recA family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF229167; AAK00736.1; -;
 DR HSSP: P26345; IG19.
 DR HAMAP: MF_00268; -; 1.
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR001553; RecA.
 DR Pfam: PF00154; RecA; 1.
 DR PRINTS: PR00142; RECA.
 DR ProDom: PD000229; RecA; 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00321; RECA_1; 1.
 DR PROSITE: PS50162; RECA_2; 1.
 DR PROSITE: PS50163; RECA_3; 1.
 DR DNA damage; DNA recombination; SOS response; ATP-binding; DNA-binding.
 FT NP_BIND 64 71
 FT VARIANT 64 71 ATP (POTENTIAL).
 SQ SEQUENCE 343 AA; 37296 MW; 06CE4CF9C1BBB48 CRC64;
 Query Match 41.4%; Score 4i: DB 1; Length 343;
 Best Local Similarity 58.3%; Pred. No. 17;
 Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 6 ERMGRGRTSSKE 17

| : : : : : : |

Db 294 ERLGGGRNSKQ 305

RESULT 9

IGF2_BOVIN ID IGF2_BOVIN STANDARD: PRI: 155 AA.
 AC P07456;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-MAY-1982 (Rel. 22, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE Insulin-like growth factor II precursor (IGF-II) (Erythrotropin) (Fragment).
 DE IGF2.
 GN IGF2.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE OF 6-155 FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=90356421; PubMed=2388846;
 RA Brown W.M., Dziegielewska K.M., Foreman R.C., Saunders N.R.;
 RT "The nucleotide and deduced amino acid sequences of insulin-like growth factor II cDNAs from adult bovine and fetal sheep liver.";
 RL Nucleic Acids Res. 18:4614-4614(1990).
 RN [2]

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: U00666; AAB60626.1; JOINED.
 DR EMBL: U00666; AAB60626.1; JOINED.
 DR EMBL: X15248; CAA33324.1; JOINED.
 DR EMBL: X53554; CAA37621.1; JOINED.
 DR EMBL: M89788; AAA31548.1; JOINED.
 DR EMBL: M89789; AAA31549.1; JOINED.
 DR EMBL: X55638; CAA39163.1; JOINED.
 DR PIR: S04858; S04858.
 DR HSSP: P01344; IGF2.
 DR InterPro: IPR004825; Ins/IGF/relax.
 DR Pfam: PF00049; Insulin; 1.
 DR PRINTS: PRO0277; INSULIN.
 DR SMART: SM00262; IIGF; 1.
 DR PROSITE: PS00262; INSULIN; 1.
 KW Insulin family; Mitogen; Growth factor; Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 91 INSULIN-LIKE GROWTH FACTOR II.
 FT DOMAIN 25 52 B.
 FT DOMAIN 53 64 C.
 FT DOMAIN 65 85 A.
 FT DOMAIN 86 91 D.
 FT PROPEP 92 179 E PEPTIDE.
 FT DISULFID 33 71 BY SIMILARITY.
 FT DISULFID 45 84 BY SIMILARITY.
 FT DISULFID 70 75 BY SIMILARITY.
 FT CONFLICT 46 47 GD -> DG (IN REF. 5).
 SQ SEQUENCE 179 AA; 19616 MW; 7B369AE57F2E4378 CRC64;

Query Match 40.4%; Score 40; DB 1; Length 179;
 Best Local Similarity 52.9%; Pred. No. 12;
 Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 VQHERMGRGRTSSKEL 18
 : I I I I I I I I I I
 DB 130 LPAFLRARRGRTLAKEL 146

RESULT 11

RECA_RHOSH STANDARD; PRT; 343 AA.
 AC P32725;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE RECA protein (Recombinase A).
 GN RECA.
 OS Rhodobacter sphaeroides (Rhodopseudomonas sphaeroides).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
 OX Rhodobacteraceae; Rhodobacter.
 OX NCBI_TaxID=1063;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-APCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
 RX MEDLINE-94104596; PubMed-8277942;
 RA Caleo S., Fernandez de Henestoga A.R., Barbe J.,
 RT "Molecular cloning, sequence and regulation of expression of the recA
 RT gene of the phototrophic bacterium Rhodobacter sphaeroides.",
 RL Mol. Gen. Genet. 242:116-120(1994).
 CC -!- FUNCTION: CAN CATALYZE THE HYDROLYSIS OF ATP IN THE PRESENCE OF
 CC SINGLE-STRANDED DNA. THE ATP-DEPENDENT UPTAKE OF SINGLE-STRANDED
 CC DNA BY DUPLEX DNA, AND THE ATP-DEPENDENT HYBRIDIZATION OF
 CC HOMOLOGOUS SINGLE-STRANDED DNAs. IT INTERACTS WITH LEXA CAUSING
 CC ITS ACTIVATION AND LEADING TO ITS AUTOCATALYTIC CLEAVAGE.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the recA family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC The European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: X72705; CAA51258.1; JOINED.
 DR PIR: S41560; S41560.
 DR HSSP: P03017; 2REB.
 DR HAMAP: MF_00268; 1.
 DR InterPro: IPR003593; AAA_ATPase.
 DR PIRfam: PF00153; RECA.
 DR Pfam: PF00154; RECA; 1.
 DR PRINTS: PRO0142; RECA.
 DR PRODOM: PD000229; RECA; 1.
 DR SMART: SMG0382; AAA; 1.
 DR PROSITE: PS00321; RECA_1; 1.
 DR PROSITE: PS0162; RECA_2; 1.
 DR PROSITE: PS0163; RECA_3; 1.
 KW DNA damage; DNA recombination; SOS response; ATP-binding; DNA-binding.
 FT NP_BIND 64 71 ATP (BY SIMILARITY).
 SQ SEQUENCE 343 AA; 36698 MW; 40584E781E5CADEF CRC64;

Query Match 40.4%; Score 40; DB 1; Length 343;
 Best Local Similarity 53.8%; Pred. No. 25;
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 4 GHERMGRGRTSSK 16
 I I I I I I I I I I
 DB 292 GDERIGGRENKAK 304

RESULT 12

CBIA_SALTY STANDARD; PRT; 459 AA.
 AC P29946;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cobyric acid A.C-diamide synthase.
 GN CBIA OR STM2035.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LT2;
 RX MEDLINE-93273696; PubMed-8501034;
 RA Roth J.R., Lawrence J.G., Rubenfield M., Kieffer-Higgins S.,
 RA Church G.M.;
 RT "Characterization of the cobalamin (vitamin B12) biosynthetic genes
 RT of Salmonella typhimurium.",
 RL J. Bacteriol. 175:3303-3316(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE-21534948; PubMed-11677609;
 RA McLelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Milvanev E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT LT2.",
 RL Nature 413:852-856(2001).
 RN [3]
 RP SEQUENCE OF 1-106 FROM N.A.
 RC STRAIN-LT2;
 RX MEDLINE-92244044; PubMed-1374146;
 RA Richter-Dahlfors A.A., Andersson D.I.;
 RT "Cobalamin (vitamin B12) repression of the Cob operon in Salmonella
 RT typhimurium requires sequences within the leader and the first

```

RT translated open reading frame."
RL Mol. Microbiol. 6:743-749(1992).
CC -!- FUNCTION: RESPONSIBLE FOR THE AMIDATION OF CARBOXYLIC GROUPS AT
CC POSITION 1 AND C OF EITHER COBIRYNIC ACID OR HYDROGENOBIRYNIC ACID.
CC NH(2) GROUPS ARE PROVIDED BY GLUTAMINE, AND ONE MOLECULE OF ATP
CC IS HYDROGENOLYZED FOR EACH AMIDATION.
CC -!- PATHWAY: Cobalam biosynthesis.
CC -!- SIMILARITY: Belongs to the cobB/cobQ family. CobB subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL: L12006; AAA27252.1; .
CC EMBL: AF008789; AAL20939.1; .
CC EMBL: X63012; CA44740.1; .
CC PIR: S20553; S20553.
CC DR StyGene: SG10034; cbIA.
CC DR HAMAP: MF_00027; . 1.
CC DR InterPro: IPR004484; CbiA.
CC DR Pfam: PF01656; CbiA; 1.
CC DR TIGRFAMs: TIGR00379; cobB; 1.
CC KW Cobalam biosynthesis; Porphyrin biosynthesis; Complete protome.
CC FT CONFLICT 36 36 R > P (IN REF. 1).
CC FT CONFLICT 111 111 M > I (IN REF. 2).
CC FT CONFLICT 128 128 V > I (IN REF. 1).
CC FT CONFLICT 133 133 A > T (IN REF. 1).
CC SQ SEQUENCE 459 AA: 50036 MW: 51615A37F5C14 CRC64:
Query Match 40.4%; Score 40; DB 1; Length 459;
Best Local Similarity 50.0%; Pred. No. 35;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 2 VPGERMGRGRT 13
DB 358 IPGSKMKRLT 169
RESULT: 13
ID REB1_SCHPO STANDARD; PRT: 504 AA.
AC Q9P6H9; O94422.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA-binding protein reb1.
GN REB1 OR SPBC1198.11C OR SPBC660.01C.
OS Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_taxid=4896;
RN [1]
RP SEQUENCE FROM N.A., AND FUNCTION.
RX MEDLINE-97169306; PubMed-9016645;
RA Zhao A., Guo A., Liu Z., Pape L.;
RT "Molecular cloning and analysis of Schizosaccharomyces pombe Reb1p:
RT sequence-specific recognition of two sites in the far upstream rDNA
RT intergenic spacer."
RL Nucleic Acids Res. 25:904-910(1997).
RN [2]
RP SEQUENCE FROM N.A.
RN STRAIN-972.
RX MEDLINE-21848401; PubMed-11859360;
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,

```

```

RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Welljans I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabe, C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Beyer K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl I.M.,
RA Hoyer P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen O., Potashkin J.,
RA Shpakovski G.V., Ossery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe."
CC -!- FUNCTION: DNA-binding protein that recognizes sites within both
CC the enhancer and the promoter of rRNA transcription, as well as
CC upstream of many genes transcribed by RNA polymerase II. Has a
CC role in the termination of RNA polymerase I catalyzed
CC transcription.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: Contains 2 Myb-like domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL: U33010; . NOT ANNOTATED_CDS.
CC EMBL: AL355921; CAB91186.1; .
CC EMBL: AL034563; CAA2522.1; .
CC GeneDB:SPombe:SUBC1198.11C; .
CC DR InterPro: IPR001005; Myb_DNA_Binding.
CC DR Pfam: PF00249; myb_DNA-binding; 2.
CC DR SMART: SM00717; SANT; 2.
CC DR PROSITE: PS00037; MYB_1; 2.
CC DR PROSITE: PS20334; MYB_2; 1.
CC DR PROSITE: PS50090; MYB_3; 2.
CC KW Nuclear protein; DNA-binding; Transcription regulation; Repeat.
FT DNA_BIND 313 357 MYB 1.
FT DNA_BIND 365 418 MYB 2.
FT SEQUENCE 504 AA: 58448 MW: 667C45F453E4F916 CRC64:
Query Match 40.4%; Score 40; DB 1; Length 504;
Best Local Similarity 37.5%; Pred. No. 39;
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
QY 1 NVPGERMGRGRTSSK 16
DB 223 NIPGYEKYSRKNSSGR 238
RESULT: 14
ID ROM_HUMAN STANDARD; PRT: 730 AA.
AC P52772;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Heterogeneous nuclear ribonucleoprotein M (hnRNP M).
GN HNRP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93181232; PubMed=8441556;
 RA Datar K.V., Dreyfuss G., Swanson M.S.;
 RI "The human hnRNP M proteins: identification of a methionine/arginine-
 rich repeat motif in ribonucleoproteins.";
 RL Nucleic Acids Res. 21:439-446(1993).
 RN [2]
 RP SEQUENCE OF 1-59 FROM N.A.
 RX MEDLINE=96292257; PubMed=8692593;
 RA Gatttoni R., Mahe C., Mahe P., Fischer N., Mattei M.-G., Stevenin C.,
 RA Fuchs J.-P.;
 RI "The human hnRNP-M proteins: structure and relation with early heat
 shock-induced splicing arrest and chromosome mapping.";
 RL Nucleic Acids Res. 24:2535-2542(1996).
 CC -!- FUNCTION: PRE-MRNA BINDING PROTEINS IN VIVO, AND THEY BIND AVIDLY
 CC TO POLY(G) AND POLY(U) RNA HOMOPOLYMERS IN VITRO. INVOLVED IN
 CC SPLICING.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Comment=Experimental confirmation may be lacking for some
 CC isoforms;
 CC Name=M4;
 CC IsoId=P52272-1; Sequence=Displayed;
 CC Name=M1-M2;
 CC IsoId=P52272-2; Sequence=VSP_005845;
 CC Name=M3;
 CC IsoId=P52272-3; Sequence=Not described;
 CC -!- SIMILARITY: Contains 3 RNA recognition motif (RRM) domains.
 CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L03532; AAA36192.1;
 DR PIR: S35532; S35532.
 DR HSSP: P1340; ICVJ.
 DR Genbank: HGNC:5046; HNRPM.
 DR GK: P52272;
 DR MIM: 160994;
 DR GO: GO:0005887; C: integral to plasma membrane; TAS.
 DR GO: GO:0005624; C: membrane fraction; TAS.
 DR InterPro: IPR000504; RNA_rec_mot.
 DR Pfam: PF00076; Rrm_3.
 DR SMART: SM00360; RRM; 3.
 DR PROSITE: PS0102; RRM; 3.
 DR PROSITE: PS0030; RRM_RNP_1; FALSE_NEG.
 KW Nuclear protein; RNA-binding; Ribonucleoprotein; Repeat;
 KW Alternative splicing.
 FT DOMAIN 71 149 RNA-BINDING (RRM) 1.
 FT DOMAIN 204 281 RNA-BINDING (RRM) 2.
 FT DOMAIN 390 396 POLY-GLY.
 FT DOMAIN 400 608 27 X 6 AA REPEATS OF G-[IL]-[DE]-R-M-
 FT [GA].
 FT REPEAT 400 405 1.
 FT REPEAT 407 412 2.
 FT REPEAT 415 420 3.
 FT REPEAT 426 431 4.
 FT REPEAT 433 438 5.
 FT REPEAT 440 445 6.
 FT REPEAT 446 451 7.
 FT REPEAT 453 458 8.
 FT REPEAT 461 466 9.
 FT REPEAT 468 473 10.
 FT REPEAT 475 480 11.
 FT REPEAT 482 487 12.

PT REPEAT 493 498 13.
 FT REPEAT 500 505 14.
 FT REPEAT 507 512 15.
 FT REPEAT 514 519 16.
 FT REPEAT 521 526 17.
 FT REPEAT 528 533 18.
 FT REPEAT 540 545 19.
 FT REPEAT 547 552 20.
 FT REPEAT 554 559 21.
 FT REPEAT 562 567 22.
 FT REPEAT 567 572 23.
 FT REPEAT 575 580 24.
 FT REPEAT 580 585 25.
 FT REPEAT 588 593 26.
 FT REPEAT 603 608 27.
 FT DOMAIN 612 616 POLY-GLY.
 FT DOMAIN 653 729 RNA-BINDING (RRM) 3.
 FT VARSPLIC 160 198 Missing (in isoform M1-M2).
 FT CONFLICT 24 34 /FTId=VSP_005845.
 FT SEQUENCE 730 AA; 77470 MW; 82236BE1D34AE27 CRC64;
 SQ
 Query Match 40.48; Score 40; DB 1; Length 730;
 Best Local Similarity 72.78; Pred. No. 58;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 Oy 1 NVPGHERMGRG 11
 Db 397 SVPGIERMGP 407
 :||| ||||| |
 RESULT 15
 POLX IOBAC STANDARD; PRT; 1328 AA.
 AC P10978;
 DT 01-JUL-1989 (Rel. 11, Created)
 DI 01-JUL-1989 (Rel. 11, Last sequence update)
 DE 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Retrovirus-related Pol polyprotein from transposon TNT 1-94 [Contains:
 DE Protease (EC 3.4.23.-); Reverse transcriptase (EC 2.7.7.49);
 DE Endonuclease].
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Lamiales; Solanales; Solanaceae; Nicotiana.
 OX NCBI_TaxID=4097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89097311; PubMed=2536143;
 RA Grandbastien M.-A., Spielmann A., Caboche M.;
 RI "Ntl, a mobile retroviral-like transposable element of tobacco
 RI isolated by plant cell genetics.";
 RL Nature 337:376-380(1989).
 CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate -> N diphosphate
 CC + (DNA)(N).
 CC -!- SIMILARITY: HIGH, WITH DROSOPHILA COP-A ELEMENT.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A11.
 CC -!- SIMILARITY: Contains 1 CCHC-type zinc finger.
 CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X13777; CAA32025.1;
 DR PIR: S04273; S04273.
 DR MEROPS: A11.002;
 DR InterPro: IPR001995; Asprotease_rtrv.
 DR InterPro: IPR001584; Rve.
 DR InterPro: IPR001878; Znf_CCHC.

DR Pfam: PF00665; rve; i.
DR Pfam: PF00077; rvp; i.
DR Pfam: PF00098; zf-CCHC; 1.
DR PRINTS: PR00939; C2HCZNFINGER.
DR SMART: SM00343; zf_C2HC; 1.
DR PROSITE: PS00158; zf_CCHC; 1.
KW Hydrolase; Aspartyl protease; RNA-directed DNA polymerase;
KW Endonuclease; transferase; Polyprotein; transposable element;
KW zinc-finger.
FT ZN_FING 230 247 CCHC-TYPE.
FT ACT_SITE 297 297 PROTEASE (BY SIMILARITY).
SQ SEQUENCE 1328 AA; 151076 MW; F27E76C504B19B1B CRC64;

Query Match 39.9%; Score 39.5; DB 1; Length 1328;
Best Local Similarity 37.5%; Pred.No. 1.4e+02;
Matches 9; Conservative 3; Mismatches 5; Indels 7; Gaps 1;

Oy 1 MVPGH-----ERMGRGRTSSKE 17
| | | | | | | | | | | | | | | |
Db 236 NQGHFKRDCNPRKKGKGTSGQK 259

Search completed: October 1, 2003, 09:38:55
Job time : 24 secs

GenCore version 5.1.6
copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2003, 09:35:49 ; Search time 3½ Seconds
(without alignments)
46.851 Million cell updates

Title: US-09-674-913A-1

Perfect score: 99
Sequence: ! NVPGHERMGRGTSSKELA 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 s

```
Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	50	50.5	72	2	A86664		hypothetical prote
2	46	46.5	342	2	A97325		homolog of cel-1 di
3	45	45.5	157	2	H82989		conserved hypothet
4	44.5	44.9	414	2	AQ1997		ribulose-bisphosph
5	43.5	43.9	498	2	T34592		DRAD-box RNA helic
6	43	43.4	79	2	F95090		conserved hypothet
7	43	43.4	79	2	B97958		conserved hypothet
8	43	43.4	151	2	S43296		bone morphogenetic
9	43	43.4	244	2	JC7801		dendritic cell-ass
10	43	43.4	424	2	T39524		hypothetical prote
11	42.5	42.9	415	2	S33627		ribulose-bisphosph
12	42.5	42.9	491	2	G95105		lysine decarboxyla
13	42.5	42.9	491	2	H97973		lysine decarboxyla
14	42	42.4	355	2	T40353		probable ribosomal
15	42	42.4	326	2	S49464		recombination prot
16	42	42.4	586	2	B87408		tetracycline resist
17	42	42.4	561	2	AE2270		exonuclease ABC c
18	42	42.4	962	2	A81817		translation initia
19	42	42.4	962	2	C81060		translation initia
20	42	42.4	1285	2	T14171		ataxin-2 - mouse
21	41.5	41.9	355	2	AC2768		alcohol dehydrogen
22	41.5	41.9	368	2	E97548		alcohol dehydrogen
23	41	41.4	235	2	T36923		hypothetical prote
24	41	41.4	302	2	AB3573		oligopeptide trans
25	41	41.4	331	2	A42602		recombination prot
26	40	40.4	69	2	AD1147		hypothetical prote
27	40	40.4	69	2	AD1506		hypothetical prote
28	40	40.4	155	1	IGBO2		insulin-like grow
29	40	40.4	162	2	F96495		probable myosin he

30	40	40.4	179	2	S04856	insulin-like growth
31	40	40.4	194	2	A83605	transcription regu
32	40	40.4	197	2	B72623	hypothetical prote
33	40	40.4	215	2	T49743	probable trm-type
34	40	40.4	240	2	H90779	DNA-binding protei
35	40	40.4	242	2	C95640	hypothetical prote
36	40	40.4	256	2	D87149	conserved hypotnet
37	40	40.4	260	2	A48488	saccharide biosynt
38	40	40.4	285	2	G83934	hypothetical prote
39	40	40.4	342	2	A22653	alcohol dehydrogen
40	40	40.4	342	2	H97434	alcohol dehydrogen
41	40	40.4	343	2	S41560	recombination prot
42	40	40.4	423	2	A70947	hypothetical prote
43	40	40.4	439	2	T02453	hypothetical prote
44	40	40.4	447	2	I22163	hypothetical prote
45	40	40.4	455	2	G70113	replicative DNA he

ALIGNMENTS

RESULT 1

A86664
hypothetical protein ydB [imported] - Lactococcus lactis subsp. lactis (Strain IL
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: A86664
R:Bolotin, A.; Wincker, P.; Mauget, S.; Jallou, O.; Malarne, K.; Weissenbach, J.;
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lac
A:Reference number: A86625; MDID:21235186; PMID:11337471
A:Accession: A86664
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-72 <STO>
A:Cross-references: GB:AEO05176; PID:g:2723578; PIDN:AAK0411.1; GSPDB:GN00146
A:Experimental source: strain IL14C3
C:Genetics:
A:Gene: ydB

Query Match	50.5%	Score 50;	DB 2;	Length 72;
Best Local Similarity	50.0%;	Pred. No. 0.35;		
Matches	9;	Conservative	3;	Mismatches 3;
				Indels 0;
				Gaps 0;

Qy	3	PCHRMGRGRTSSKE	17
		1111111111	
Qp	46	PIDHEKMGKG7TLSBE	60

RESULT 2

A:7325
homolog of cell division GTPase FtsZ, diverged [imported]; - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence-revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: A97325
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacteri
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: A97325
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-342 <KUR>
A:Cross-references: GB:AE001437; PIDW:AAK81388.1; PID:gl5026550; GSPDB:GN00158
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC3459

Query Match	46.5%	Score 46;	DB 2;	Length: 342;
Best Local Similarity	56.2%	Pred. No. 7.6;		
Matches	9;	Conservative	2;	Mismatches 5;
				Indels 0;
				Gaps 0;

GenCore version 5.1.6					
Copyright (c) 1993 - 2003 Compugen Ltd.					
OM protein - protein search, using sw model					
Run on:	October 1, 2003, 09:35:49 ; Search time 3 1/2 seconds (without alignments) 46.851 Million cell updates/sec				
Title:	US-09-674-913A-1				
Perfect score:	99				
Sequence:	1 NVPCHERMGRGRTSSKELA 19				
Scoring table:	BLOSUM62 Gapop 10.0 , Gapext 0.5				
Searched:	283308 seqs, 96168682 residues				
Total number of hits satisfying chosen parameters:	283308				
Minimum DB seq length:	0				
Maximum DB seq length:	2000000000				
Post-processing:	Minimum Match 0% Maximum Match 100% Listing first 45 summaries				
Database :	PIR_75:* 1: pir1:* 2: pir2:* 3: pir3:* 4: pir4:*				
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.					
SUMMARIES					
Result No.	Score	Query Match	Length	DB ID	Description
1	50	50.5	72	2 A86664	hypothetical prote
2	46	46.5	342	2 A97325	homolog of cell di
3	45	45.5	157	2 H82989	conserved hypothet
4	44.5	44.9	414	2 AG1837	ribulose-bisphosph
5	43.5	43.9	498	2 T34592	DEAD-box RNA helic
6	43	43.4	75	2 F95090	conserved hypothet
7	43	43.4	79	2 B97958	conserved hypothet
8	43	43.4	151	2 S43296	bone morphogenetic
9	43	43.4	244	2 JC7801	dendritic cell-ass
10	43	43.4	424	2 T39524	hypothetical prote
11	42.5	42.9	415	2 S33627	ribulose-bisphosph
12	42.5	42.9	491	2 G95105	lysine decarboxyla
13	42.5	42.9	491	2 H97973	lysine decarboxyla
14	42	42.4	326	2 T40353	probable ribosomal
15	42	42.4	355	2 S49464	recombination prot
16	42	42.4	586	2 B87408	tetracycline resis
17	42	42.4	961	2 AE2270	exonuclease ABC c
18	42	42.4	962	2 A81817	translation initia
19	42	42.4	962	2 C81060	translation initia
20	42	42.4	1285	2 I14171	ataxin-2 - mouse
21	41.5	41.9	355	2 AC2768	alcohol dehydrogen
22	41.5	41.9	368	2 E97548	alcohol dehydrogen
23	41	41.4	235	2 T36923	hypothetical prote
24	41	41.4	302	2 T36923	hypothetical prote
25	41	41.4	331	2 A32602	oligopeptide trans
26	40	40.4	69	2 AD3147	recombination prot
27	40	40.4	69	2 AD1506	hypothetical prote
28	40	40.4	155	1 IGB021	insulin-like grow
29	40	40.4	162	2 F96495	probable myosin he

```
QY 3 PHERMGRGRTSSKEL 18
      || : || || |||
Db 57 PGIDSGGRDR*KSXEM 72

RESULT 3
H82989
conserved hypothetical protein PA5246 [imported] - Pseudomonas aeruginosa (strain: PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H82989
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; R-
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: AB2950; MUID:2043737; PMID:10984043
A:Accession: H82989
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-157 <STO>
A:Cross-references: GB:AE004937; GB:AE004091; NID:g9951553; PIDN:AA0C8631.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA5246

Query Match 45.5%; Score 45; DB 2; Length 157;
Best Local Similarity 36.7%; Pred. No. 5.1;
Matches 11; Conservative 1; Mismatches 0; Indels 18; Gaps 1;

QY 2 VPGHERMGRGRTSSKEL 13
      |||||
Db 89 LPGHERMARLSKLGIDLRVDYLRPGRT 118

RESULT 4
AG1997
ribulose-bisphosphate carboxylase activase (EC 6.3.4.-) [similarity] - Nostoc sp. (strain
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AG1997
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Katanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG1997
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-414 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA077899.1; PID:g17135353; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: rca
C:Superfamily: ribulose-bisphosphate carboxylase activase
C:Keywords: ATP; ligase; nucleotide binding; P-loop
F:37-44/Region: nucleotide-binding motif A (P-loop)
F:43/Binding site: ATP (Lys) #status predicted

Query Match 44.9%; Score 44.5; DB 2; Length 414;
Best Local Similarity 39.3%; Pred. No. 16;
Matches 11; Conservative 2; Mismatches 6; Indels 9; Gaps 1;

QY 1 NVPGHERMGRGRTSSKELA 19
      |||
Db 23 NIFGVRVPLILGHRKGEKIFQCELA 50

RESULT 5
T34592
DEAD-box RNA helicase - streptomyces coelicolor
C:Species: Streptomyces coelicolor

C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 17-Nov-2000
C:Accession: T34592
R:Murphy, L.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, January 1998
A:Reference number: 221548
A:Accession: T34592
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-498 <SMUR>
A:Cross-references: EMBL:AL021529; PIDN:CAAL6457.1; GSPDB:GN00070; SCOEDB:SC10A5.25c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC10A5.25c

Query Match 43.9%; Score 43.5; DB 2; Length 498;
Best Local Similarity 55.0%; Pred. No. 29;
Matches 11; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

QY 1 NYPGHERMGRGRTSSKELA 19
      || : ||||| ||| ||
Db 103 SLAGRDIMGRGRTSGKTLA 122

RESULT 6
F95090
conserved hypothetical protein SP0782 [imported] - Streptococcus pneumoniae (strain
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: F95090
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.;
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapf
nson, T.; Hickey, E.K.; Holt, I.E.
Science 291, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venier, J.C.; Dougherty, B.A.; Morr
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: F95090
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-79 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK74919.1; PID:g14972257; GSPDB:GN00164; TIG
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP0782

Query Match 43.4%; Score 43; DB 2; Length 79;
Best Local Similarity 53.3%; Pred. No. 5.5;
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 PHERMGRGRTSSKE 17
      || : ||| |||
Db 53 PDHTRMKRGRTISNE 67

RESULT 7
H97958
conserved hypothetical protein spr0630 [imported] - Streptococcus pneumoniae (strain
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C:Accession: H97958
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.;
O., K.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAnren, S
y, P.; Sun, P.M.; Winkler, M.F.;
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: H97958
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-79 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99494.1; PID:g15458279; GSPDB:GN00174
C:Genetics:
```


A:Gene: spr0690

Query Match 43.4%; Score 43; DB 2; Length 79;

Best Local Similarity 53.3%; Pred. No. 5.5;

Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 PCHRMGRGRSSKE 17

|||||

DB 53 PDHTRMGKGTLSNF 67

RESULT 8

S43296

bone morphogenetic protein-related protein (GDF7) - mouse

C:Species: Mus musculus (house mouse)

C>Date: 20-Oct-1994 #sequence_revision 10-Nov-1995 #text_change 19-May-2000

C:Accession: S43296

R:Storm, E.E.; Huynh, T.V.; Cope land, N.G.; Jenkins, K.A.; Kingsley, D.M.; Lee, S.J.

Nature 368, 639-643, 1994

A:Title: Limb alterations in brachypodism mice due to mutations in a new member of the T

A:Reference number: S43294; MUID:94195427; PMID:8145850

A:Accession: S43296

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <STO>

A:Cross-References: G3:U08339; NID:9488465; PIDN:AAAL9780.1; PID:q488465

C:Superfamily: inhibitor

Query Match 43.4%; Score 43; DB 2; Length 151;

Best Local Similarity 60.0%; Pred. No. 10;

Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 GHERMGRGRTSSKEL 18

|||||

DB 41 GHRGRGRGRTSSKSL 55

RESULT 9

JC7801

dendritic cell-associated nuclear protein 1, DCNP1 - human

C:Species: Homo sapiens (man)

C>Date: 03-Jun-2002 #sequence_revision 03-Jun-2002 #text_change 03-Jun-2002

C:Accession: JC7801

R:Masuda, M.; Senju, S.; Fujii, S.; Terasaki, Y.; Takeya, M.; Hashimoto, S.; Matsushima,

Biochem. Biophys. Res. Commun. 290, 1022-1029, 2002

A:Title: Identification and immunocytochemical analysis of DCNP1, a dendritic cell-associated

A:Reference number: JC7801; PMID:11798177; MUID:21656978

A:Accession: JC7801

A:Molecule type: mRNA

A:Residues: 1-244 <MAS>

A:Cross-References: DDBJ:AB074498

C:Comment: This protein is a dendritic cell-associated nuclear membrane protein involved

C:Genetics:

A:Gene: dcnp1

A:Map position: 5

Query Match

Best Local Similarity 43.4%; Score 43; DB 2; Length 244;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 VPGHMRMG 11

|||||

DB 20 VPGHQRLEGR 29

RESULT 10

T39524

hypothetical protein SPBC665.08 - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T39524

R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Hilbert, H.; Duesterhoeft, A.

submitted to the EMBL Data Library, July 1998

A:Reference number: 221861

A:Accession: T39524

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-424 <MOO>

A:Cross-References: EMBL:AL031154; PIDN:CAA20056.1; GSPDB:GN00067; SPDB:SPBC1695.0

A:Experimental source: strain 972h-; cosmid c1685

C:Genetics:

A:Gene: SPDB:SPBC1685.08

A:Map position: 2

A:Insertions: 113/2; 161/2

Query Match

Best Local Similarity 43.4%; Score 43; DB 2; Length 424;

Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 7 RMGRGRGRTSSKELA 19

|||||

DB 177 RRRGRGRTSSKSL 189

RESULT 11

S33627

ribulose-bisphosphate carboxylase activase (EC 6.3.4.-) - Anabaena sp. (strain CA)

C:Species: Anabaena sp.

C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Jan-2001

C:Accession: S33627

P:Li, L.A.; Gibson, J.L.; Tabita, F.R.

Plant Mol. Biol. 21, 753-764, 1993

A:Title: The rubisco activase (rca) gene is located downstream from rbcS in Anabaena

A:Reference number: S33627; MUID:93222475; PMID:8467074

A:Accession: S33627

A:Molecule type: DNA

A:Residues: 1-415 <LIL>

A:Cross-References: EMBL:X67942; NID:q296413; PIDN:CAA48129.1; PID:q296414

C:Genetics:

A:Gene: rca

C:Superfamily: ribulose-bisphosphate carboxylase activase

C:Keywords: ATP; ligase; nucleotide binding; P-loop

F:37-44/Region: nucleotide-binding motif A (P-loop)

F:43/Binding site: ATP (Lys) #status predicted

Query Match

Best Local Similarity 42.9%; Score 42.5; DB 2; Length 415;

Matches 11; Conservative 2; Mismatches 6; Indels 9; Gaps 1;

QY 1 NVPG-----HERMGRGRTSSKELA 19

|||||

DB 23 NLPGVRVPLILGIRGKGKGTQCELA 50

RESULT 12

G95105

lysine decarboxylase [imported] - Streptococcus pneumoniae (strain TIGR4)

C:Species: Streptococcus pneumoniae

C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001

C:Accession: G95105

R:Tettelein, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.

on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtza

son, T.; Hickey, E.K.; Holt, I.E.

Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Mor

A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: G95105

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-491 <KUR>

A:Cross-References: G3:AE005672; PIDN:AAK75040.1; PID:q14972390; GSPDB:GN00164; T1

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP0916

C:Superfamily: lysine decarboxylase cad

Query Match 42.9%; Score 42.5; DB 2; Length 491;

Best Local Similarity 55.6%; Pred. No. 41;

Matches 10; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

QY 1 NVPGHERMGRGRTSSKEL 18

||||| |||

Db 27 DVPGHKR-GRGNPELVEL 43

RESULT 13

H97973 lysine decarboxylase (EC 4.1.1.18) [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae

C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001

C:Accession: H97973

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burgett, S.; DeHoff, B.S.; B

e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M

y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;

A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: H97973

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-491 <KUR>

A:Cross-references: GB:AE007317; PIDN:AAK99620.1; PID:g15458416; GSPDB:GN00174

C:Genetics:

A:Gene: cad

C:Superfamily: lysine decarboxylase cad

C:Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 42.9%; Score 42.5; DB 2; Length 491;

Best Local Similarity 55.6%; Pred. No. 41;

Matches 10; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

QY 1 NVPGHERMGRGRTSSKEL 18

||||| |||

Db 27 DVPGHKR-GRGNPELVEL 43

RESULT 14

I40353

probable ribosomal protein - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T40353

R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Skelton, J.; Churcher, C.M.

submitted to the EMBL Data Library, March 1997

A:Reference number: Z21922

A:Accession: T40353

A>Status: preliminary; translated from GB/EMBL/DDHJ

A:Molecule type: DNA

A:Residues: 1-326 <WOC>

A:Cross-references: EMBL:AL020270; PIDN:CAA17794.1; GSPDB:GN00067; SPDB:SPBC3B9.14C

A:Experimental source: strain 972h-; cosmid c3B9

C:Genetics:

A:Gene: SPDB:SPBC3B9.14C

A:Map position: 2

A:Introns: 188/3; 275/3

Query Match 42.4%; Score 42; DB 2; Length 326;

Best Local Similarity 50.0%; Pred. No. 33;

Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 4 GHERMGRGRTSSKEIA 19

||| ||| ||

Db 257 GHLLGGQASSLNLA 272

RESULT 15

S49464

recombination protein recA - Rhodobacter capsulatus

N:Alternate names: recombinase A

C:Species: Rhodobacter capsulatus

C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 02-Feb-2001

C:Accession: S49464

R:Fernandez de Henestrosa, A.R.

submitted to the EMBL Data Library, October 1994

A:Reference number: S49464

A:Accession: S49464

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-355 <FER>

A:Cross-references: EMBL:X82183; NID:g558630; PIDN:CAA57673.1; PID:g558631

C:Genetics:

A:Gene: recA

C:Superfamily: recombination protein recA

C:Keywords: ATP; DNA binding; DNA recombination; DNA repair; nucleotide binding; P-

F:78-85/Region: nucleotide-binding motif A (P-loop)

F:152-157/Region: nucleotide-binding motif B

F:84/Binding site: ATP (Lys) #status predicted

Query Match 42.4%; Score 42; DB 2; Length 355;

Best Local Similarity 50.0%; Pred. No. 36;

Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 4 GHERMGRGRTSSKE 17

||| ||| |||

Db 306 GDERIGQGRENKQ 319

Search completed: October 1, 2003, 09:41:26

Job time : 42 secs